New Solution for H-infinity Static Output-Feedback Control Using Integral Reinforcement Learning

Yusuf Kartal[a], Wenqian Xue[b], Ahmet Taha Koru[a], Frank L. Lewis[a], Atilla Dogan[a]

[a] *University of Texas at Arlington, UTA Research Institute, Fort Worth, 76118, TX, USA*
[b] *Northeastern University, State Key Laboratory of Synthetical Automation for Process Industries and International Joint Research Laboratory of Integrated Automation, Shenyang, 110819, China*

Abstract

This paper presents a new formulation for the H_∞ static output-feedback (OPFB) control problem that guarantees stability, L_2 gain boundedness, and optimal stabilizing gain solutions for a Linear Time-Invariant (LTI) system. Then, based on the developed formulation, it reveals an integral reinforcement learning algorithm (IRL) that employs Kleinman's method and achieves stabilizing optimal control strategies without requiring any knowledge of the system state, control, and disturbance matrices. The standard approaches to the H_∞ static OPFB control problem result in sub-optimal and non-Nash gain solutions for guaranteed stability. They propose an offline algorithm that may converge to only a sub-optimal stabilizing gain solution. On the other hand, this paper proposes a novel augmented Hamiltonian functional to solve the global Nash equilibrium solutions for a game of this kind. Based on the augmented Hamiltonian's stationarity conditions, we provide novel necessary and sufficient conditions for Nash equilibrium gain solutions that inherently stabilize the system dynamics while also guaranteeing L_2 gain bound by a prescribed attenuation level. To obtain the Nash solution without knowledge of system parameters, two off-policy IRL algorithms are developed based on Kleinman's algorithm. In the first IRL algorithm, the convergence to the Nash gain solution is provided assuming system state data is available. Then, a second novel IRL algorithm is developed that does not require system state data and learns the Nash equilibrium gain solution online. Simulation results are provided to show the validity of the proposed methods.

Preprint submitted to Systems and Control Letters *August 6, 2022*

Keywords: H_∞ optimal control, integral reinforcement learning, zero-sum game, bounded L_2 gain

1. Introduction

One of the primary objectives in control system design is often to seek a stabilizing controller to regulate the output of a system that experiences disturbances. However, stability is not the only requirement in control system design. An L_2 gain bound of the system, optimality of the control method, and detectability of unknown system parameters are other common design specifications. Existing solutions of H_∞ static output-feedback (OPFB) yield stability and bounded L_2 gain, but employee non-Nash equilibrium solutions. In the two-player zero-sum game context, the Nash equilibrium consists of optimal strategies for both players. Based on this, the new formulation of H_∞ static OPFB control method is developed in this paper, which is a key to meet these requirements since it guarantees L_2 gain bound of the system by a prescribed attenuation level, asymptotic stability of equilibrium point, and also Nash equilibrium solutions. Then, based on the new formulae, the IRL algorithm is developed that iteratively solves H_∞ static OPFB Lyapunov equation, and deals with unknown system parameters.

H_∞ control methods has been widely studied in the literature, [1], [2], [3], [4], [5], [6] due to their applicability in variety of engineering areas. Some of these methods guarantee stability and bounded L_2 gain, but an extra condition is required to yield Nash equilibrium. [1] uses this method to design a gain-scheduled normal acceleration control loop for an air-launched unmanned aerial vehicle. Authors of [2] apply this control method on an industrial-type mass spring damper system. The efficacy of control law and the disturbance accommodation properties are shown on a rotor-craft design example in [7]. Moreover, [8] develop an autopilot controller for an F-16 aircraft by using the H_∞ static OPFB control method on a linear discrete-time system.

During the last few years, reinforcement learning (RL) algorithms [9], [10] has been used extensively to replace model parameters with a collected system's data [11], [12], [13], [14], [15], [16]. Particularly, offline iterative RL algorithms were studied in [17], [18], and [19]. The work by [17] considers the two-player policy iterations to solve for the feedback strategies of a continuous-time zero-sum game [20] in a sub-optimal manner that requires complete knowledge of system parameters. [12] presented an online RL

algorithm to solve the linear quadratic tracking (LQT) problem for partially-unknown continuous-time systems. In [21], the authors prove the convergence of IRL algorithm to a sub-optimal OPFB solution without considering the disturbance term when the drift dynamics are unknown. The optimal average cost learning framework is introduced to solve the output regulation problem for linear systems with unknown dynamics is studied in [22].

To design an efficient RL algorithm and achieve data-driven optimal control, the RL-based controller designs with neural networks (NNs) in an actor-critic structure [23], critic-only form [24] are proposed. In the off-policy RL algorithm, the system data, which are used to learn the solution of the corresponding Hamiltonian, can be generated with arbitrary policies rather than the evaluating policy. This approach is suitably implemented using NNs in [25] and [13].

The standard existing solutions to the static OPFB regulator problem [1], [21], [26], [27], [28] require some additive gain matrix to prove the stability of equilibrium, origin. Unfortunately, this results in the non-Nash equilibrium solutions. Consequently, the IRL algorithms developed based on these approaches [28], result in sub-optimal, non-Nash solutions. In this paper, we propose a novel augmented Hamiltonian, and develop new iterative algorithms based on stationarity conditions of the augmented Hamiltonian to obtain Nash equilibrium solutions. The salient contributions of this paper are summed up into four categories:

- This paper presents a new solution of the H_∞ static OPFB control problem. This solution guarantees not only stability and bounded L_2 gain but also Nash equilibrium solutions considering the corresponding min-max game.

- The existing methods employ an offline algorithm that does not have a convergence guarantee. Further, if the algorithm converges, it only converges sub-optimal stabilizing gain solutions. Our new solution rigorously yields Nash gain solutions given the novel necessary and sufficient conditions.

- To develop an IRL algorithm based on the augmented Hamiltonian's stationarity conditions, two off-line iterative solution algorithms are given. The first algorithm is based on Kleinman's algorithm and updates the disturbance gain term. A second algorithm modifies Kleinman's algorithm and gets the first algorithm in the IRL applicable form

that only updates the control gain.

- Two off-policy IRL algorithms are developed based on the modified Kleinman's algorithm. The first IRL algorithm learns the Nash gain solution online without requiring any knowledge of system dynamics' state, control, and disturbance matrices assuming the system state date is available. The second IRL algorithm assumes only the output data is available and develops a model-free observer to learn the Nash gain solution.

The rest of the paper is organized as follows. In Section 2, preliminaries on control design requirements are introduced, and the formulation of H_∞ static OPFB control is presented. A new solution of optimal H_∞ control problem and corresponding offline iterative solution algorithms are given in Section 3. In Section 4, an online off-policy IRL algorithm is developed based on stationarity conditions obtained in Section 3. Finally, we have shown the effectiveness of the proposed algorithms by applying them to the linearized lateral dynamics of the F-16 aircraft at a particular flight condition in Section 5.

Notations. We use the following notations throughout this paper $\boldsymbol{I_n} \in \mathbb{R}^{n\times n}$ is the identity matrix. The condition $A > 0$ (≥ 0) denotes the positive (semi) definiteness of a matrix. The operator $tr()$ denotes trace of a matrix. $\boldsymbol{C^+ = C^T(CC^T)^{-1}}$ is the right-inverse of the full row-rank matrix $\boldsymbol{C}$ and the Kronecker product operator is denoted by $\otimes$. The determinant of a square matrix is denoted by $|.|$. $vec(A)$ stands for the mn-vector formed by stacking the columns of $A \in \mathbb{R}^{n\times m}$ on top of one another, i.e., $vec(A) = [a_1^T \ ... \ a_m^T]^T$ where $a_i \in \mathbb{R}^n$ are the columns of A. Lastly, $diag(\zeta_i)$ represents a diagonal matrix with $\zeta_i \ \forall i \in 1, ..., N$ on its diagonal.

2. Preliminaries and problem formulation

In this section, preliminaries on Linear Time Invariant (LTI) system, and the corresponding controller design requirements are first introduced. Then, the problem description is presented.

2.1. System description and definitions

This section introduces system dynamics and performance specifications that are of interest. Consider the state-space representation of the continuous-

time LTI system as

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{D}\boldsymbol{d} \\ \boldsymbol{y} &= \boldsymbol{C}\boldsymbol{x} \end{aligned} \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times m}$, $\boldsymbol{D} \in \mathbb{R}^{n\times p}$ are system-state, input, disturbance matrices, and $\boldsymbol{C} \in \mathbb{R}^{q\times n}$ is assumed to be a full row-rank output matrix to avoid redundant measurements. The corresponding vectors $\boldsymbol{x}(t)$, $\boldsymbol{u}(t)$, $\boldsymbol{d}(t)$, and $\boldsymbol{y}(t)$ stand for the state, input, disturbance and output respectively.

Assumption 1. The pair $(\boldsymbol{A}, \boldsymbol{B})$ is *stabilizable* and the pair $(\boldsymbol{A}, \boldsymbol{C})$ is *detectable*.

Assumption 2. The system (1) is OPFB stabilizable because the row-space of output matrix $\boldsymbol{C}$ contains the sub-space spanned by the right eigenvectors corresponding to the unstable modes of $\boldsymbol{A}$.

Assumption 3. The non-zero columns of the output matrix $\boldsymbol{C}$ are linearly independent.

Remark 1. The Assumption 2 can be interpreted such that all unstable modes are measured by the output matrix $\boldsymbol{C}$ that represents the sensors installed in the system (1). The Assumption 3 enables us to recover a state element x_i precisely from the output vector $\boldsymbol{y}$ once it is left multiplied with C^+.

Now, define the fictitious performance output $\boldsymbol{z}(t)$ that satisfies

$$\|\boldsymbol{z}\|_2^2 = \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{u}^{\boldsymbol{T}}\boldsymbol{R}\boldsymbol{u} \tag{2}$$

where $\boldsymbol{Q} \geq 0$ and $\boldsymbol{R} > 0$ are symmetric design matrices with appropriate dimensions. We assume that $\boldsymbol{Q}$ is selected such that the pair $(\boldsymbol{A}, \sqrt{\boldsymbol{Q}})$ is *observable*, which is a standard assumption [29]. Using the property $\|\boldsymbol{d}\|_2^2 = \boldsymbol{d}^{\boldsymbol{T}}\boldsymbol{d}$, a realization of the following inequality $\forall \boldsymbol{d} \in [0, \infty)$ implies that the system L_2-gain is bounded by a prescribed disturbance attenuation level denoted by γ

$$\int_0^\infty \|\boldsymbol{z}\|_2^2 dt \leq \gamma^2 \int_0^\infty \|\boldsymbol{d}\|_2^2 dt + \beta \tag{3}$$

for any non-zero energy-bounded disturbance input d [30] where β is a non-negative constant. The condition (3) is also called as non-expansivity constraint in [31]. Call γ^* the minimum gain for which this occurs. In [32] and

[33], an algorithm to find γ^* is given, and a formulation for explicit γ^* that depends on Riccati equation solution is derived for LTI systems under some assumptions. This paper assumes that the attenuation level is prescribed and satisfies $\gamma > \gamma^*$.

A static OPFB control to regulate the system (1) is

$$\boldsymbol{u} = -\boldsymbol{K}\boldsymbol{y} = -\boldsymbol{K}\boldsymbol{C}\boldsymbol{x} \tag{4}$$

where $\boldsymbol{K} \in \mathbb{R}^{m\times q}$ is the gain matrix. Note that main objective of H_∞ control using OPFB is to find the stabilizing $\boldsymbol{K}$ in an optimal manner while satisfying the condition (3), which can be achieved by solving corresponding Hamilton-Jacobi-Isaacs (HJI) equation.

2.2. Problem formulation and existing solution of static OPFB Problem

In this section, we relate zero-sum differential game theory to the static OPFB regulation problem in a global optimal manner by revealing various definitions. To satisfy L_2-gain bound (3) with the stabilizing gain in (4), an objective functional defined as

$$J(\boldsymbol{u}, \boldsymbol{d}) = \int_0^\infty (\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{u}^{\boldsymbol{T}}\boldsymbol{R}\boldsymbol{u} - \gamma^2\boldsymbol{d}^{\boldsymbol{T}}\boldsymbol{d})d\tau. \tag{5}$$

Now H_∞ control problem can be represented as a two-player zero-sum differential game by treating $\boldsymbol{u}(t)$ as a minimizing player, whereas $\boldsymbol{d}(t)$ maximizing player of (5). Then, the game can be formulated as

$$V(\boldsymbol{x}(\boldsymbol{0})) = J(\boldsymbol{u}^*, \boldsymbol{d}^*) = \min_{\boldsymbol{u}} \max_{\boldsymbol{d}} J(\boldsymbol{u}, \boldsymbol{d}) \tag{6}$$

where $V(\boldsymbol{x})$ denotes the value functional corresponding to (5) such that

$$V(\boldsymbol{x}) = \int_t^\infty (\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{u}^{\boldsymbol{T}}\boldsymbol{R}\boldsymbol{u} - \gamma^2\boldsymbol{d}^{\boldsymbol{T}}\boldsymbol{d})d\tau. \tag{7}$$

and the pair $(\boldsymbol{u}^*, \boldsymbol{d}^*)$ denotes the game theoretic saddle point. The game of this kind admits a unique solution pair $(\boldsymbol{u}^*, \boldsymbol{d}^*)$, if the following Nash condition holds

$$\min_{\boldsymbol{u}} \max_{\boldsymbol{d}} J(\boldsymbol{u}, \boldsymbol{d}) = \max_{\boldsymbol{d}} \min_{\boldsymbol{u}} J(\boldsymbol{u}, \boldsymbol{d}). \tag{8}$$

The next theorem recalls the necessary and sufficient conditions for the sub-optimal H_∞ OPFB control method [1].

Theorem 1. *The system* (1) *is OPFB stable using the control* $\boldsymbol{u_o^e = -K_o^e y}$ *with* L_2 *gain bounded by* $\gamma > \gamma^*$ *if and only if*

1. $(\boldsymbol{A}, \boldsymbol{B})$ *is stabilizable and* $(\boldsymbol{A}, \boldsymbol{C})$ *is detectable.*

2. *There exists* $\boldsymbol{K_o^e}$ *and* $\boldsymbol{L}$ *such that*

$$\boldsymbol{K_o^e = R^{-1}(B^T P + L)C^+} \tag{9}$$

where $\boldsymbol{P = P^T > 0}$ *is the solution of Riccati equation*

$$\boldsymbol{PA + A^T P - PBR^{-1}B^T P + Q + \gamma^{-2} PDD^T P + L^T R^{-1} L = 0}. \tag{10}$$

Proof. See [1] and [30] for the same proof. □

Remark 2. On the one hand, introducing the additive gain matrix $\boldsymbol{L}$ provides the extra design freedom. Note that if $\boldsymbol{L = 0}$ there may not be a stabilizing solution to (9),(10) (See Theorem 1 in [1]). On the other hand, this results in a sub-optimal solution for the gain $\boldsymbol{K^s}$ (9). The Nash equilibrium gain occurs if the Theorem 1 holds with $\boldsymbol{L = 0}$.

The following lemma recalls the Nash equilibrium solution for the standard H_∞ control problem.

Lemma 1. The pair $(\boldsymbol{u}^*, \boldsymbol{d}^*)$ constitutes the Nash equilibrium of the game (8) such that

$$\boldsymbol{u^* = -K^* x = -R^{-1} B^T P x}, \tag{11}$$

$$\boldsymbol{d^* = \gamma^{-2} D^T P x}. \tag{12}$$

Proof. Begin with deriving the Hamiltonian to solve the game theoretic saddle point or Nash equilibrium strategy of the game (8) as

$$H(\boldsymbol{V_x, u, d}) = \boldsymbol{x^T Q x + u^T R u - \gamma^2 d^T d + V_x(Ax + Bu + Dd)} \tag{13}$$

where $\boldsymbol{V_x} = \partial V / \partial \boldsymbol{x}$ is the co-state vector. Using the quadratic form $V(\boldsymbol{x}) = \boldsymbol{x^T P x}$, and applying the stationarity conditions $\partial H(\boldsymbol{V_x, u, d})/\partial \boldsymbol{u} = \boldsymbol{0}$ and $\partial H(\boldsymbol{V_x, u, d})/\partial \boldsymbol{d} = \boldsymbol{0}$ yields the optimal control and disturbance respectively as (11) and (12).

Notice that the sign of Hessians, $\partial^2 H(\boldsymbol{V_x}, \boldsymbol{u}, \boldsymbol{d})/\partial \boldsymbol{u^2} > 0$ and $\partial^2 H(\boldsymbol{V_x}, \boldsymbol{u}, \boldsymbol{d})/\partial \boldsymbol{d^2} < 0$, along with unboundedness of the limits $\lim_{d\to\infty} J(\boldsymbol{u}^*, \boldsymbol{d})$, $\lim_{u\to\infty} J(\boldsymbol{u}, \boldsymbol{d}^*)$ indeed show that (11) and (12) are the global optimal minimizing and maximizing extrema respectively [34]. This indeed verifies that the pair $(\boldsymbol{u}^*, \boldsymbol{d}^*)$ denotes the Nash equilibrium point, which completes the proof. □

Remark 3. The HJI equation, $H(\boldsymbol{V_x}, \boldsymbol{u}^*, \boldsymbol{d}^*) = 0$ with the boundary condition $V(0) = 0$ can be obtained by substituting the expressions (11)-(12) into Hamiltonian (13), which also verifies the sub-optimality of gain expression (9). Additionally, from the HJI equation $H(\boldsymbol{V_x}, \boldsymbol{u}^*, \boldsymbol{d}^*) = 0$, the Game Algebraic Riccati Equation (GARE) can be obtained as

$$\boldsymbol{PA} + \boldsymbol{A^T P} - \boldsymbol{PBR^{-1}B^T P} + \boldsymbol{Q} + \gamma^{-2}\boldsymbol{PDD^T P} = \boldsymbol{0}. \tag{14}$$

If there is no $\boldsymbol{K^s}$ to satisfy (9) and (10), the OPFB H_∞ control problem may not have even a sub-optimal solution. In the next section, we rigorously analyze this and reveal some novel results.

3. New solution of $\boldsymbol{H_\infty}$ OPFB Game

Notice that Theorem 1 provides necessary and sufficient conditions for static OPFB in a sub-optimal manner. Thence, this does not yield a Nash equilibrium solution unless $L = 0$. Moreover, there may not even exist a static OPFB solution to the equations in Theorem 1. In this main section, two methods are proposed to solve H_∞ OPFB problem. The first method parameterizes the state feedback gains by using the Nash strategies and applies them to the OPFB design. The second method derives the new optimal H_∞ OPFB regulator formulation by introducing an augmented Hamiltonian. This method is introduced by [29], but is highly overlooked in the literature. However, it appears to be instrumental in H_∞ regulator design.

3.1. Necessary and Sufficient Conditions for the Stabilizing Nash Gain

Herein, we first parameterize static state feedback gains and then explain how to apply them to the H_∞ OPFB design.

Theorem 2. *Given the necessary conditions in the Assumption 1 and the sufficient condition* $\boldsymbol{BR^{-1}B^T} \geq \gamma^{-2}\boldsymbol{DD^T}$. *The system* (1) *is asymptotically stable using the control* $\boldsymbol{u^*} = -\boldsymbol{R^{-1}B^T Px}$ (11) *with* $d = 0$, *and* L_2 *gain bounded by* $\gamma\ \forall \|\boldsymbol{d}\|_2 \in (0, \infty)$.

Proof. To prove L_2 gain bound condition (3), first re-write the Hamiltonian (13) by completing the squares as

$$\begin{aligned} H(\boldsymbol{V_x}, \boldsymbol{u}, \boldsymbol{d}) =& H(\boldsymbol{V_x}, \boldsymbol{u^*}, \boldsymbol{d^*}) + (\boldsymbol{u} - \boldsymbol{u}^*)^T \boldsymbol{R}(\boldsymbol{u} - \boldsymbol{u}^*) \\ & - \gamma^{-2}(\boldsymbol{d} - \boldsymbol{d}^*)^T(\boldsymbol{d} - \boldsymbol{d}^*). \end{aligned} \tag{15}$$

Then, the objective functional can be re-expressed as

$$\begin{aligned} J(\boldsymbol{u}, \boldsymbol{d}) = \int_0^\infty \Big(& H(\boldsymbol{V_x}, \boldsymbol{u^*}, \boldsymbol{d^*}) + (\boldsymbol{u} - \boldsymbol{u}^*)^T \boldsymbol{R}(\boldsymbol{u} - \boldsymbol{u}^*) \\ & - \gamma^{-2}(\boldsymbol{d} - \boldsymbol{d}^*)^T(\boldsymbol{d} - \boldsymbol{d}^*) \Big) dt + V(\boldsymbol{x}(0)). \end{aligned} \tag{16}$$

Realize that the non-expansivity constraint (3) implies that $J(\boldsymbol{u}, \boldsymbol{d}) \leq \beta$. Select $\beta = V(\boldsymbol{x}(0))$, $\boldsymbol{u} = \boldsymbol{u}^*$, and note that HJI equation $H(\boldsymbol{V_x}, \boldsymbol{u^*}, \boldsymbol{d^*}) = 0$ holds with the boundary condition V(0)=0. Then, (16) reduces to

$$\begin{aligned} J(\boldsymbol{u^*}, \boldsymbol{d}) &= -\int_0^\infty \gamma^{-2}(\boldsymbol{d} - \boldsymbol{d}^*)^T(\boldsymbol{d} - \boldsymbol{d}^*) dt + V(\boldsymbol{x}(0)), \\ &\leq V(\boldsymbol{x}(0)), \quad \forall \left\| \boldsymbol{d} \right\|_2 \in (0, \infty). \end{aligned} \tag{17}$$

This proves that the L_2 gain bound condition (3) holds with $\beta = V(\boldsymbol{x}(0))$, $\boldsymbol{u} = \boldsymbol{u}^*$. Additionally, the value of game (8) with $\boldsymbol{u} = \boldsymbol{u}^*$ and $\boldsymbol{d} = \boldsymbol{d}^*$ is $V(\boldsymbol{x}(0))$ by (16).

To verify asymptotic stability, consider Lyapunov function $V(\boldsymbol{x}) = \boldsymbol{x^T P x}$ where $\boldsymbol{P}$ is solution of the GARE (14). Note that for $\boldsymbol{B R^{-1} B^T} \geq \gamma^{-2} \boldsymbol{D D^T}$, the GARE (14) has a unique stabilizing solution $\boldsymbol{P} = \boldsymbol{P^T}$, which is indeed positive definite. This is illustrated in the following realization

$$\begin{aligned} \dot{V} &= \dot{\boldsymbol{x}}^{\boldsymbol{T}} \boldsymbol{P x} + \boldsymbol{x^T P} \dot{\boldsymbol{x}} \\ &= \boldsymbol{x^T} \big(- \boldsymbol{P B R^{-1} B^T P} - \boldsymbol{Q} + \gamma^{-2} \boldsymbol{P D D^T P} \big) \boldsymbol{x} \\ &\leq -\boldsymbol{x^T Q x} \quad \Leftarrow \ \boldsymbol{B R^{-1} B^T} \geq \gamma^{-2} \boldsymbol{D D^T}. \end{aligned} \tag{18}$$

Then, the observability of (A, Q) verifies that the undisturbed system (1), i.e., $d = 0$, is asymptotically stable by LaSalle's invariance principle [31]. This completes the proof. □

Corollary 1. To verify asymptotic stability of the disturbed system, benefit from gain margin $[c_{lower}, \infty)$ with $c_{lower} < \frac{1}{2}$ property of the $\mathcal{H}_\infty$ control [31].

Note that the $\mathcal{H}_\infty$ has gain margin less than $\frac{1}{2}$ by Chapter 10 in [31] but the lower bound c_{lower} is not precisely defined. Then, if the sufficient condition in Theorem 2 is strengthened as $\boldsymbol{BR^{-1}B^T} \geq 2\gamma^{-2}\boldsymbol{DD^T}$, the disturbed system (1) with $\boldsymbol{d} = \boldsymbol{d}^*$, becomes asymptotically stable. Thence, the closed-loop matrix $\boldsymbol{A - BR^{-1}B^TP + \gamma^{-2}DD^TP}$ becomes Hurwitz.

Till now, we actually parameterize all the stabilizing static state-feedback gains since we used the Nash strategies (11) and (12) in the proof of Theorem 2. Note that for the OPFB design, the Assumptions 2 and 3 are missing in the papers [1], [2], [7], and [30]. Therefore, the offline algorithm given in these works do not have a guaranteed convergence. In this manuscript, we use Assumptions 2 and 3 to apply state feedback gains to the OPFB design. The following Remark explains how to apply static state feedback gains to the H_∞ OPFB design.

Remark 4. Instead of Nash strategies (11) and (12), assume that the control and disturbance are selected as

$$\boldsymbol{u_o^*} = -(\boldsymbol{R^{-1}B^TPC^+})\boldsymbol{C}(\boldsymbol{C^+y}) = -\underbrace{(\boldsymbol{R^{-1}B^TPC^+})}_{\boldsymbol{K_o^*}}\boldsymbol{y} \tag{19}$$

$$\boldsymbol{d_o^*} = (\gamma^{-2}\boldsymbol{D^TPC^+})\boldsymbol{C}(\boldsymbol{C^+y}). \tag{20}$$

Note that if $\boldsymbol{C}$ is an invertible matrix, then all states would be regulated optimally by Theorem 2. On the other hand, if it is not square but full row-rank, then only the states spanned by row space of the output matrix $\boldsymbol{C}$ would be regulated optimally given the Assumption 3, and the fact that C^+C projects $\mathbb{R}^n$ onto the row space of $\boldsymbol{C}$. Additionally, the other states would converge to the origin given the Assumption 2. Realize that the Assumption 1 implies that there could be an unstable mode that is observable but does not belong to the row space of $\boldsymbol{C}$. Therefore, the Assumptions 2 and 3 are indeed required to apply static state feedback gains to the OPFB design. Lastly, the system (1) is stable against the worst-case disturbance (20), which affects only the states that belong to the row space of $\boldsymbol{C}$ if $\boldsymbol{BR^{-1}B^T} \geq 2\gamma^{-2}\boldsymbol{DD^T}$ by Corollary 1.

3.2. A Direct Method to Obtain OPFB Optimal Gain Solutions

In this section, we propose a new methodology to obtain stabilizing Nash solutions for the H_∞ OPFB control. This method is direct in the sense that

it reaches the same gain solutions as the Section 3.1 but do not require two step

Consider the optimal value obtained in the Theorem 2 that corresponds to the zero-effort for each player such that

$$J_0 = \boldsymbol{x}^{\boldsymbol{T}}(0)\boldsymbol{P}\boldsymbol{x}(0) = tr(\boldsymbol{P}\boldsymbol{X}_{\boldsymbol{0}}) \tag{21}$$

where $tr()$ stands for the trace of a matrix, and $\boldsymbol{X}_{\boldsymbol{0}} = \boldsymbol{x}(0)\boldsymbol{x}^{\boldsymbol{T}}(0)$.

The following Lemma is an essential step before introducing the augmented Hamiltonian.

Lemma 2. Given the Assumption 1, let $\boldsymbol{K}^{\boldsymbol{a}}$ be a gain that stabilizes the system (1), and the corresponding OPFB control is $\boldsymbol{u}_{\boldsymbol{o}}^{\boldsymbol{a}} = -\boldsymbol{K}_{\boldsymbol{o}}^{\boldsymbol{a}}\boldsymbol{y}$. Additionally, let the disturbance takes the form $\boldsymbol{d}_{\boldsymbol{o}}^{\boldsymbol{a}} = -\boldsymbol{N}_{\boldsymbol{o}}\boldsymbol{y}$ to guarantee it does not affect unobservable modes of the system (1). Then the corresponding Lyapunov equation can be derived as

$$\boldsymbol{P}\boldsymbol{A}_{\boldsymbol{c}} + \boldsymbol{A}_{\boldsymbol{c}}^{\boldsymbol{T}}\boldsymbol{P} + \boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{K}_{\boldsymbol{o}}^{\boldsymbol{a}\boldsymbol{T}}\boldsymbol{R}\boldsymbol{K}_{\boldsymbol{o}}^{\boldsymbol{a}}\boldsymbol{C} - \gamma^2\boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{N}_{\boldsymbol{o}}^{\boldsymbol{T}}\boldsymbol{N}_{\boldsymbol{o}}\boldsymbol{C} + \boldsymbol{Q} \equiv \boldsymbol{T} \tag{22}$$

where $\boldsymbol{T} = \boldsymbol{T}^{\boldsymbol{T}} = \boldsymbol{0}$ and $\boldsymbol{A}_{\boldsymbol{c}} = \boldsymbol{A} - \boldsymbol{B}\boldsymbol{K}_{\boldsymbol{o}}^{\boldsymbol{a}}\boldsymbol{C} + \boldsymbol{D}\boldsymbol{N}_{\boldsymbol{o}}\boldsymbol{C}$.

Proof. Consider the quadratic form of the value functional (5) as $V(\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{x}$, and then substitute expressions $\boldsymbol{u} = -\boldsymbol{K}^{\boldsymbol{a}}\boldsymbol{y}$ and $\boldsymbol{d}_{\boldsymbol{o}}^{\boldsymbol{a}} = -\boldsymbol{N}_{\boldsymbol{o}}\boldsymbol{y}$ into (5) to obtain

$$\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{x} = \int_t^{\infty} \boldsymbol{x}^{\boldsymbol{T}}(\boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{K}_{\boldsymbol{o}}^{\boldsymbol{a}\boldsymbol{T}}\boldsymbol{R}\boldsymbol{K}_{\boldsymbol{o}}^{\boldsymbol{a}}\boldsymbol{C} + \boldsymbol{Q} - \gamma^2\boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{N}_{\boldsymbol{o}}^{\boldsymbol{T}}\boldsymbol{N}_{\boldsymbol{o}}\boldsymbol{C})\boldsymbol{x}d\tau. \tag{23}$$

Now, take the derivative of left-side (23) and substitute (12), $\boldsymbol{u} = -\boldsymbol{K}^{\boldsymbol{a}}\boldsymbol{C}\boldsymbol{x}$ expressions. Lastly, take the derivative of integral in right-side (23) using Leibniz's rule, which yields

$$\begin{aligned}\boldsymbol{x}^{\boldsymbol{T}}(\boldsymbol{A}_{\boldsymbol{c}}^{\boldsymbol{T}}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}_{\boldsymbol{c}})\boldsymbol{x} = \\ \boldsymbol{x}^{\boldsymbol{T}}(-\boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{K}_{\boldsymbol{o}}^{\boldsymbol{a}\boldsymbol{T}}\boldsymbol{R}\boldsymbol{K}_{\boldsymbol{o}}^{\boldsymbol{a}}\boldsymbol{C} - \boldsymbol{Q} + \gamma^2\boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{N}_{\boldsymbol{o}}^{\boldsymbol{T}}\boldsymbol{N}_{\boldsymbol{o}}\boldsymbol{C})\boldsymbol{x}.\end{aligned} \tag{24}$$

Realize that the zero equivalent is nothing but the Lyapunov equation given in (22). This completes the proof. □

It is now clear that performing a min-max operation on (5), subject to dynamical constraint (1) is equivalent to the algebraic problem of finding the pair $(\boldsymbol{K}_{\boldsymbol{o}}^{\boldsymbol{a}}, \boldsymbol{N}_{\boldsymbol{o}})$ that performs min-max of (21) subject to the constraint (22).

Define the following augmented Hamiltonian to solve this modified problem as

$$H^a(\boldsymbol{K_o^a}, \boldsymbol{N_o}, S) = tr(\boldsymbol{P X_0}) + tr(\boldsymbol{T S}) \tag{25}$$

where $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ is a symmetric matrix of Lagrange multipliers [29] that needs to be determined, and $\boldsymbol{T}$ is given in (22). Notice that $\boldsymbol{T}$ includes weighting matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$.

The next main theorem is a key to find $\boldsymbol{S}$ along with a Nash equilibrium control matrix $\boldsymbol{K_o^a}$ with respect to (25).

Theorem 3. *Given the Assumptions 1-3, the system* (1) *is asymptotically stable using the control* $\boldsymbol{u_o^a} = -\boldsymbol{K_o^a y}$ *with* $d_o^a = 0$ *and* L_2 *gain bounded by* γ *if*

$$\frac{\partial H^a}{\partial \boldsymbol{S}} \equiv \boldsymbol{P A_c} + \boldsymbol{A_c^T P} + \boldsymbol{C^T K_o^{aT} R K_o^a C} - \gamma^2 \boldsymbol{C^T N_o^T N_o C} + \boldsymbol{Q} = \boldsymbol{0}, \tag{26}$$

$$\frac{\partial H^a}{\partial \boldsymbol{P}} \equiv \boldsymbol{S A_c^T} + \boldsymbol{A_c S} + \boldsymbol{X_0} = \boldsymbol{0}, \tag{27}$$

$$\frac{\partial H^a}{\partial \boldsymbol{K_o^a}} \equiv 2\boldsymbol{R K_o^a C S C^T} - 2\boldsymbol{B^T P S C^T} = \boldsymbol{0}, \tag{28}$$

$$\frac{\partial H^a}{\partial \boldsymbol{N_o}} \equiv -2\gamma^2 \boldsymbol{N_o^a C S C^T} + 2\boldsymbol{D^T P S C^T} = \boldsymbol{0}. \tag{29}$$

Furthermore, the following gain expressions solves the H_∞ *static OPFB problem in an optimal manner*

$$\boldsymbol{K_o^a} = \boldsymbol{K_o^*} = \boldsymbol{R^{-1} B^T P C^+} \tag{30}$$

$$\boldsymbol{N_o} = \gamma^{-2} \boldsymbol{D^T P C^+} \tag{31}$$

Proof. Consider (25), which is a constant along the system trajectories since the system (1) is LTI and $\boldsymbol{z}(t)$ in (2) is not explicit function of time. This implies that we can apply the constraint test and check the stationarity conditions on the augmented Hamiltonian (25) that yields the second-order Lyapunov equation (26) and standard Lyapunov equation (27) respectively.

Define a variable $\boldsymbol{X} = \boldsymbol{x x^T}$ that includes the system state information. Taking the derivative of $\boldsymbol{X}$ using (1) yields

$$\begin{aligned} \dot{\boldsymbol{X}} &= \dot{\boldsymbol{x}} \boldsymbol{x^T} + \boldsymbol{x} \dot{\boldsymbol{x}}^{\boldsymbol{T}} \\ &= \boldsymbol{A_c x x^T} + \boldsymbol{x x^T A_c^T} \\ &= \boldsymbol{A_c X} + \boldsymbol{X A_c^T}. \end{aligned} \tag{32}$$

Now, assume that $\boldsymbol{A_c}$ is Hurwitz. Then, taking the integral of both sides (32) from 0 to ∞ yields (27) where $S = \int_0^\infty X dt$, thereby $\boldsymbol{K_o^a}$ should not depend on the solution S. Therefore, the gain solutions (30) and (31) are immediate. Realize that the stability of an LTI system does not depend on the initial condition, i.e, local stability implies global stability. Thence, the gain solutions (30) and (31) should not depend on S that depends on the initial condition X_0. This also verifies our reason to select gain solutions in the given forms, which completes the proof. □

Remark 5. The Theorem 3 gives the necessary conditions, i.e. the Assumption 1 and 2, and sufficient condition $\boldsymbol{BR^{-1}B^T} \geq \gamma^{-2}\boldsymbol{DD^T}$ to prove L_2 gain boundedness by a prescribed attenuation level γ (3) and OPFB stability considering the worst-case disturbance (20). Realize that the condition $\boldsymbol{BR^{-1}B^T} \geq \gamma^{-2}\boldsymbol{DD^T}$ is only a sufficient condition, which implies that there may be an optimal gain solution which stabilizes (1) but does not satisfy $\boldsymbol{BR^{-1}B^T} \geq \gamma^{-2}\boldsymbol{DD^T}$. However, in that case, one may not achieve a positive definite solution P for the Riccati equation (10). Additionally, the positive definiteness of the Riccati equation solution plays a key role for Kleinman's algorithm in Section 3.3, and the IRL algorithm in Section 4.

Remark 6. The Theorem 3 proves that the condition $\boldsymbol{BR^{-1}B^T} \geq \gamma^{-2}\boldsymbol{DD^T}$ is $sufficient$ to obtain the stabilizing Nash equilibrium solution. The gain $\boldsymbol{K_o^e}$ in Theorem 1 is a sub-optimal stabilizing gain solution with respect to the value functional (7). However, the gain solution K^a in Theorem 3 always gives a stabilizing Nash equilibrium gain solutions with respect to the game (8) given the Assumptions 1-3.

Remark 7. Note that the system (1) is stable in the presence of matched disturbances with the gains (30) and (31). The unmatched disturbances are only L_2 gain bounded by Theorem 2. Thence, the control (19) is robustly stabilizing the equilibrium origin even if the unmatched disturbances exist given Assumptions 2 and 3.

The next section proposes an offline model-based algorithm to find the optimal gain solution $\boldsymbol{K_o^a}$ iteratively, that plays a key role to develop the IRL Algorithm that will be detailed later in Section 4. Note that given the necessary and sufficient conditions in Theorem 3 and Remark 5, one does not need an iterative solution algorithm to find optimal stabilizing gain $\boldsymbol{K_o^a}$ (30). However, to develop a model-free algorithm, an iterative solution algorithm is required.

3.3. Offline Iterative Solution Algorithms for H_∞ OPFB

This section presents two iterative solution algorithms to obtain minimizing gain (30) by using the conditions given in (26)-(28). In the first algorithm, we employ Kleinman's algorithm (26) to obtain Nash equilibrium gain solutions (30) and (31), whereas in the second algorithm, we use a corresponding Lyapunov equation to not deal with the disturbance gain term $\boldsymbol{N_o}$.

The next algorithm performs a sequence of four-step iterations based on the Kleinman's Algorithm [35] to find the optimal control gains (30) and (31).

Algorithm 1. (Offline iterative solution with Lyapunov equations. Kleinman's Algorithm.)

1. Initialize: Set $k = 1$, $\boldsymbol{P_0} = \boldsymbol{0}$, $\boldsymbol{N_0} = \boldsymbol{0}$ and given the Assumption 1 and the sufficient condition $\boldsymbol{BR^{-1}B^T} \geq 2\gamma^{-2}\boldsymbol{DD^T}$, select a gain $\boldsymbol{F_0}$ such that $\boldsymbol{A} - \boldsymbol{BF_0}$ is asymptotically stable.

2. k^{th} iteration: Solve for $\boldsymbol{P_k}$

$$\boldsymbol{0} = \boldsymbol{P_k A_k} + \boldsymbol{A_k^T P_k} + \boldsymbol{F_{k-1}^T R F_{k-1}} + \boldsymbol{Q} \tag{33}$$

   where $\boldsymbol{A_k} = \boldsymbol{A} - \boldsymbol{BF_k} + \frac{1}{2}DN_k$. Finally, update the gains

$$\boldsymbol{F_k} = \boldsymbol{R^{-1}B^T P_k}, \tag{34}$$

$$\boldsymbol{N_k} = \gamma^{-2}\boldsymbol{D^T P_k}. \tag{35}$$

   Set the cost $J_k = tr(\boldsymbol{P_k X_0})$.

3. Check: If F_{k-1} and F_k are close enough to each other, go to step (4) else go to step (2).

4. Terminate: Given the Assumptions 2 and 3, set the OPFB gains $\boldsymbol{K_o^a} = \boldsymbol{F_k C^+}$, $\boldsymbol{N_o} = \boldsymbol{N_k C^+}$ and the cost $\boldsymbol{J} = \boldsymbol{J_k}$. □

Note that the closed-loop stability, and L_2 gain boundedness implies that (36) has a unique stabilizing optimal solution $\boldsymbol{P} > \boldsymbol{0}$ by Theorem 2. A comprehensive study for the solution of generalized Riccati equations can be found in [36]. The Algorithm 2 is based on the iterative solution algorithm presented in [37] whose convergence is proved by establishing the connection

between Newton's method in [38]. Additionally, by considering the condition $\boldsymbol{BR^{-1}B^T} \geq \gamma^{-2}\boldsymbol{DD^T}$, the monotonic convergence, i.e, $P_k < P_{k-1}$, is straight-forward from the Theorem 13.5.8 in [38]. A related algorithm is also examined in Chapter 8 of the book [29].

Now, to develop an algorithm that accounts only for the optimal gain K_o^a, we manipulate the steps of the Algorithm 1. The resultant Algorithm 2 finds the the optimal control gain K_o^a (30) iteratively.

Algorithm 2. (Offline iterative solution with Lyapunov equations. Modified Kleinman's Algorithm.)

1. Initialize: Set $k = 1$, $\boldsymbol{P_0} = \boldsymbol{0}$, and given the Assumption 1 and the sufficient condition $\boldsymbol{BR^{-1}B^T} \geq 2\gamma^{-2}\boldsymbol{DD^T}$, select a gain $\boldsymbol{F_0}$ such that $\boldsymbol{A} - \boldsymbol{BF_0}$ is asymptotically stable.

2. k^{th} iteration: Solve for $\boldsymbol{P_k}$

$$\boldsymbol{0} = \boldsymbol{P_k A_k} + \boldsymbol{A_k^T P_k} + \boldsymbol{F_{k-1}^T R F_{k-1}} + \boldsymbol{Q} + \gamma^{-2}\boldsymbol{P_{k-1} D D^T P_{k-1}} \quad (36)$$

   where $\boldsymbol{A_k} = \boldsymbol{A} - \boldsymbol{BF_k}$. Finally, update the gain

$$\boldsymbol{F_k} = \boldsymbol{R^{-1}B^T P_k}. \quad (37)$$

   Set the cost $J_k = tr(\boldsymbol{P_k X_0})$.

3. Check: If F_{k-1} and F_k are close enough to each other, go to step (4) else go to step (2).

4. Terminate: Given the Assumptions 2 and 3, set the OPFB gain $\boldsymbol{K_o^a} = \boldsymbol{F_k C^+}$ and the cost $\boldsymbol{J} = \boldsymbol{J_k}$. □

The next section uses the Algorithm 2 to develop model-free algorithms considering different scenarios for the availability of system data.

4. Online Integral Reinforcement Learning Solution Algorithm for $\boldsymbol{H_\infty}$ OPFB

In this section, we first develop an online off-policy integral reinforcement learning (IRL) algorithm [39], which is a model-free version of the Algorithm 2. This algorithm assumes that the system state data is available to deal with unknown $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{D}$ matrices. Then, we develop a novel IRL algorithm that solves the optimal H_∞ regulator problem by learning the Nash equilibrium gain solution (30) without requiring knowledge of the system state data.

4.1. Online off-policy IRL algorithms

This section introduces an off-policy IRL algorithm to deal with the unknown system matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{D}$. In this case, both of Algorithm 1 and 2 lose their applicability as they are model-based. However, we still benefit from the convergence properties of Algorithm 2 while developing the off-policy IRL method. To this end, we represent system dynamics (1) as

$$\begin{aligned}\dot{\boldsymbol{x}} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}_{\boldsymbol{k}} + \boldsymbol{D}\boldsymbol{d} + \boldsymbol{B}(\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{k}}) \\ &= \boldsymbol{A}_{\boldsymbol{k}}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{d} + \boldsymbol{B}(\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{k}})\end{aligned} \tag{38}$$

where $\boldsymbol{A}_{\boldsymbol{k}} = \boldsymbol{A} - \boldsymbol{B}\boldsymbol{F}_{\boldsymbol{k}}$ and $\boldsymbol{u}_{\boldsymbol{k}} = -\boldsymbol{F}_{\boldsymbol{k}}\boldsymbol{x} \in \mathbb{R}^m$ is the control policy to be updated with $\boldsymbol{F}_{\boldsymbol{k}}$ given in Algorithm 2.

Firstly, to obtain $\boldsymbol{P}_{\boldsymbol{k}}$ without information of the system matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{D}$, take the derivative of value functional $V(\boldsymbol{x}(t)) = \boldsymbol{x}^{\boldsymbol{T}}(t)\boldsymbol{P}\boldsymbol{x}(t)$ by using the new representation of the system dynamics (38)

$$\begin{aligned}\dot{V} &= \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{A}_{\boldsymbol{k}}^{\boldsymbol{T}}\boldsymbol{P}_{\boldsymbol{k}}\boldsymbol{x} + \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{P}_{\boldsymbol{k}}\boldsymbol{A}_{\boldsymbol{k}}\boldsymbol{x} \\ &\quad + 2\boldsymbol{d}^{\boldsymbol{T}}\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{P}_{\boldsymbol{k}}\boldsymbol{x} + 2(\boldsymbol{u} + \boldsymbol{F}_{\boldsymbol{k}}\boldsymbol{x})^T\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{P}_{\boldsymbol{k}}\boldsymbol{x}.\end{aligned} \tag{39}$$

To employ the approach given in Algorithm [35], we define the following two new variables

$$\boldsymbol{G}_{\boldsymbol{k+1}} = \gamma^{-2}\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{P}_{\boldsymbol{k}}, \quad \boldsymbol{F}_{\boldsymbol{k+1}} = \boldsymbol{R}^{-\boldsymbol{1}}\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{P}_{\boldsymbol{k}}. \tag{40}$$

Re-write (36) in terms of new variable (40) to get the Algorithm 2 in the Kleinman's form as

$$\bar{\boldsymbol{Q}} = \boldsymbol{P}_{\boldsymbol{k}}\boldsymbol{A}_{\boldsymbol{k}} + \boldsymbol{A}_{\boldsymbol{k}}^{\boldsymbol{T}}\boldsymbol{P}_{\boldsymbol{k}} \tag{41}$$

where $\bar{\boldsymbol{Q}} = -\boldsymbol{F}_{\boldsymbol{k}}^{\boldsymbol{T}}\boldsymbol{R}\boldsymbol{F}_{\boldsymbol{k}} - \boldsymbol{Q} - \gamma^2\boldsymbol{G}_{\boldsymbol{k}}^{\boldsymbol{T}}\boldsymbol{G}_{\boldsymbol{k}}$. Additionally, express (39) in terms of the new variables introduced in (40) and (41) as

$$\dot{V} = 2(\gamma^2\boldsymbol{d}^{\boldsymbol{T}}\boldsymbol{G}_{\boldsymbol{k+1}} + (\boldsymbol{u} + \boldsymbol{F}_{\boldsymbol{k}}\boldsymbol{x})^T\boldsymbol{R}\boldsymbol{F}_{\boldsymbol{k+1}})\boldsymbol{x} + \boldsymbol{x}^{\boldsymbol{T}}\bar{\boldsymbol{Q}}\boldsymbol{x}. \tag{42}$$

Then, integrate both sides from t to $t+T$ to obtain

$$\begin{aligned}V(t+T) - V(t) &= \int_t^{t+T} 2(\boldsymbol{u} + \boldsymbol{F}_{\boldsymbol{k}}\boldsymbol{x})^T\boldsymbol{R}\boldsymbol{F}_{\boldsymbol{k+1}}\boldsymbol{x}d\tau \\ &\quad + \int_t^{t+T} 2\gamma^2\boldsymbol{d}^{\boldsymbol{T}}\boldsymbol{G}_{\boldsymbol{k+1}}\boldsymbol{x}d\tau + \int_t^{t+T} \boldsymbol{x}^{\boldsymbol{T}}\bar{\boldsymbol{Q}}\boldsymbol{x}d\tau.\end{aligned} \tag{43}$$

Based on these manipulations, the online off-policy IRL algorithm can be developed. Note that this new IRL algorithm and the Algorithm 2 are equivalent. However, on the contrary offline Algorithm 2, the IRL Algorithm 3 does not require information of $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{D}$ matrices. It only requires an initial stabilizing control policy $u_0 = -F_0 x$ (this is standard assumption in IRL applications [14],[23],[25],[28]) and the sufficient amount of data that belongs $\boldsymbol{x}(t)$, $\boldsymbol{u}(t)$, and $\boldsymbol{d}(t)$ vectors, which can be collected online.

Algorithm 3. (Off-policy IRL algorithm assuming $\boldsymbol{x}$ is given.)

1. Initialize: Set $k = 1$, $\boldsymbol{G_0} = \boldsymbol{0}$ and given Assumption 1, start with initial stabilizing control policy $u_0 = -F_0 x$.

2. k^{th} iteration: Use (43) to update $\boldsymbol{P_k}$, $\boldsymbol{G_{k+1}}$ and $\boldsymbol{F_{k+1}}$ simultaneously

$$\boldsymbol{x^T}(t+T)\boldsymbol{P_k}\boldsymbol{x}(t+T) - \boldsymbol{x^T}(t)\boldsymbol{P_k}\boldsymbol{x}(t) - \int_t^{t+T} 2\gamma^2 \boldsymbol{d^T}\boldsymbol{G_{k+1}}\boldsymbol{x}d\tau$$
$$- \int_t^{t+T} 2(\boldsymbol{u} + \boldsymbol{F_k}\boldsymbol{x})^T \boldsymbol{R}\boldsymbol{F_{k+1}}\boldsymbol{x}d\tau = \int_t^{t+T} \boldsymbol{x^T}\bar{\boldsymbol{Q}}\boldsymbol{x}d\tau. \tag{44}$$

   Set the cost $J_k = tr(\boldsymbol{P_k}\boldsymbol{X_0})$.

3. Check: If $\boldsymbol{F_k}$ and $\boldsymbol{F_{k+1}}$ are close enough to each other, go to step (4) else go to step (2).

4. Terminate: Given the Assumptions 2 and 3, set $\boldsymbol{K_o^a} = \boldsymbol{F_{k+1}}\boldsymbol{C^+}$ and $\boldsymbol{J} = \boldsymbol{J_k}$. □

Remark 8. Note that right-side of the equation (44) in the Algorithm 3 consists of known terms, and hence it can be solved for $\boldsymbol{P_k}$, $\boldsymbol{G_{k+1}}$ and $\boldsymbol{F_{k+1}}$ matrices using well established least-squares technique by converting them to the set of linear equations [25]. Since (44) does not use any system matrix information except $\boldsymbol{C}$, the Algorithm 3 is said to be model-free. Realize that the output matrix $\boldsymbol{C}$ represents the sensors placed to the system (1), which is clearly known.

Realize that the Algorithm 3 achieves the static OPFB gain assuming system state information, $\boldsymbol{x}$, is available. On the other hand, if we are only given the output data $\boldsymbol{y}$, we need to come up with a novel algorithm that does not require system state data. One approach is to make use of the

observability matrix while developing a model-free algorithm. However, this approach indeed requires the pair $(\boldsymbol{A}, \boldsymbol{C})$ to be observable. Thence, from now on we assume that the detectability condition in the Assumption 1 is strengthened as an observability condition.

The next Theorem is instrumental before we develop a new Algorithm that only need measurement of the system output data $\boldsymbol{y}$.

Theorem 4. *For any given n-dimensional observable system, there exists a sufficiently small time interval* $[0\ T]$ *such that if* N *sampling times satisfy* $0 \leq t - i\Delta t \leq T\ \forall i \in 1, ..., N$ *where* Δt *is the delayed time and assumed fixed, then the system is* N*-sample observable.*

Proof.: See [40] for the same proof.

To make use of the observability matrix define

$$\boldsymbol{x}(t)^T \boldsymbol{P} \boldsymbol{x}(t) = \boldsymbol{Y}^T(t) \tilde{\boldsymbol{P}} \boldsymbol{Y}(t) \tag{45}$$

where $\boldsymbol{Y}(t) = [\boldsymbol{y}^{\boldsymbol{T}}(t)\ \dot{\boldsymbol{y}}^{\boldsymbol{T}}(t)\ ...\ \boldsymbol{y}^{n-1^T}(t)]^T = \boldsymbol{O}\boldsymbol{x}$, and hence $\tilde{\boldsymbol{P}} = \mathcal{O}_L^T \boldsymbol{P} \mathcal{O}_L$ with $\mathcal{O}_L \in \mathbb{R}^{n \times nq}$ is the left-inverse of the observability matrix $\boldsymbol{O} \in \mathbb{R}^{nq \times n}$. Note that each derivative of the output $\boldsymbol{y}$ can be obtained by making use of the Taylor Series expansion on the collected data $\boldsymbol{y}(t + i\Delta t)$ around $\boldsymbol{y}(t)$. An example is $\ddot{\boldsymbol{y}} = \frac{\boldsymbol{y}(t+\Delta t) + \boldsymbol{y}(t-\Delta t) - 2\boldsymbol{y}(t)}{\Delta t^2}$ where $\boldsymbol{y}(t + \Delta t) = \boldsymbol{y}(t) + \Delta t \dot{\boldsymbol{y}}(t) + 0.5\Delta^2 t \ddot{\boldsymbol{y}}(t)$ and $\boldsymbol{y}(t - \Delta t) = \boldsymbol{y}(t) - \Delta t \dot{\boldsymbol{y}}(t) + 0.5\Delta^2 t \ddot{\boldsymbol{y}}(t)$.

Now, select $\boldsymbol{Q} = k\boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{C}$ where $k > 0$ is a scalar, and define the following variables

$$\tilde{\boldsymbol{F}}_k = \boldsymbol{F}_{\boldsymbol{k}} \mathcal{O}_L,\ \tilde{\boldsymbol{G}}_k = \boldsymbol{G}_{\boldsymbol{k}} \mathcal{O}_L, \tag{46}$$

and the known term $\tilde{\boldsymbol{Q}} = -\tilde{\boldsymbol{F}}_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{R} \tilde{\boldsymbol{F}}_{\boldsymbol{k}} - \hat{\boldsymbol{Q}} - \gamma^2 \tilde{\boldsymbol{G}}_{\boldsymbol{k}}^{\boldsymbol{T}} \tilde{\boldsymbol{G}}_{\boldsymbol{k}}$. Herein the new weighting matrix selected in a form such that $\hat{\boldsymbol{Q}} = k[\boldsymbol{I_q}\ \boldsymbol{0}\ ...\ \boldsymbol{0}]_{qn \times q}^T [\boldsymbol{I_q}\ \boldsymbol{0}\ ...\ \boldsymbol{0}]_{q \times qn}$. Realize that $\boldsymbol{x}^{\boldsymbol{T}} \boldsymbol{Q} \boldsymbol{x} = \boldsymbol{y}^{\boldsymbol{T}} \boldsymbol{y} = \boldsymbol{Y}^{\boldsymbol{T}} \hat{\boldsymbol{Q}} \boldsymbol{Y}$ holds when $\boldsymbol{Q} = k\boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{C}$, and it enables us to treat $\tilde{\boldsymbol{Q}}$ as a known term. Further, since the pair $(\boldsymbol{A}, \boldsymbol{C})$ is observable $(\boldsymbol{A}, k\boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{C})$ is also observable. Based on these manipulations, a new Algorithm 4 can be developed.

Algorithm 4. (Off-policy IRL algorithm, $\boldsymbol{x}$ is not required but the pair $(\boldsymbol{A}, \boldsymbol{C})$ must be observable.)

1. Initialize: Set $k = 1$, $\tilde{\boldsymbol{G}}_{\boldsymbol{0}} = \boldsymbol{0}$ and start with a stabilizing control policy $\boldsymbol{u} = -\tilde{\boldsymbol{F}}_{\boldsymbol{0}} \boldsymbol{Y}$.

2. k^{th} iteration: Update $\tilde{\boldsymbol{P}}_{\boldsymbol{k}}$, $\tilde{\boldsymbol{G}}_{\boldsymbol{k+1}}$ and $\tilde{\boldsymbol{F}}_{\boldsymbol{k+1}}$ simultaneously

$$\boldsymbol{Y^T}(t+T)\tilde{\boldsymbol{P}}_{\boldsymbol{k}}\boldsymbol{Y}(t+T) - \boldsymbol{Y^T}(t)\tilde{\boldsymbol{P}}_{\boldsymbol{k}}\boldsymbol{Y}(t) - \int_t^{t+T} 2\gamma^2 \boldsymbol{d^T}\tilde{\boldsymbol{G}}_{\boldsymbol{k+1}}\boldsymbol{Y}\,d\tau$$
$$- \int_t^{t+T} 2(\boldsymbol{u} + \tilde{\boldsymbol{F}}_{\boldsymbol{k}}\boldsymbol{Y})^T \boldsymbol{R}\tilde{\boldsymbol{F}}_{\boldsymbol{k+1}}\boldsymbol{Y}\,d\tau = \int_t^{t+T} \boldsymbol{Y^T}\tilde{\boldsymbol{Q}}\boldsymbol{Y}\,d\tau. \tag{47}$$

Set the cost $J_k = tr(\tilde{\boldsymbol{P}}_{\boldsymbol{k}}\boldsymbol{X_0})$.

3. Check: If $\tilde{\boldsymbol{F}}_k$ and $\tilde{\boldsymbol{F}}_{k+1}$ are close enough to each other, go to step (4) else go to step (2).

4. Terminate: Given the Assumptions 2 and 3, set $\boldsymbol{u} = -\tilde{\boldsymbol{F}}_{\boldsymbol{k+1}}\boldsymbol{Y}$ and $\boldsymbol{J} = \boldsymbol{J}_{\boldsymbol{k}}$. □

Realize that the Algorithm 4 converges with the static state feedback expression since $\boldsymbol{u} = -\tilde{\boldsymbol{F}}_{\boldsymbol{k+1}}\boldsymbol{Y} = -\boldsymbol{R^{-1}}\boldsymbol{B^T}\boldsymbol{P}_k\mathcal{O}_L\mathcal{O}\boldsymbol{x} = -\boldsymbol{R^{-1}}\boldsymbol{B^T}\boldsymbol{P}_k\boldsymbol{x}$ (11). Thence, it regulates not only the state elements spanned in the row-space of $\boldsymbol{C}$ but all states. However, it creates additional complexity as it requires more data to be collected to converge. On the other hand, the Algorithm 4 does not require system state data $\boldsymbol{x}$, and it directly achieves the Nash equilibrium control u by making use of the new variables $\tilde{\boldsymbol{F}}_{\boldsymbol{k+1}}$ and $\boldsymbol{y}$ instead of calculating the OPFB gain $\boldsymbol{K}_{\boldsymbol{o}}^{\boldsymbol{a}}$. Additionally, since the Algorithm 4 is obtained with the change of variables in the Algorithm 3, it shares similar convergence properties with the Algorithm 3, and hence the Algorithm 2. The next section shows a way of solving coupled linear equations.

4.2. Data-based implementation of the IRL Algorithm

This section introduces a least-squares method to solve step (2) in Algorithm 3. Although value function approximation is a popular tool and can be employed to solve step 2 in Algorithm 3, it requires three Neural Networks (NNs), i.e., the actor NN to approximate the value functional (7), the critic NN to update control policy and the disturber NN to update disturbance policy [13]. This causes a complicated NN design procedure. Instead, we use a least-squares method to solve for $\boldsymbol{P}_{\boldsymbol{k}}$ in step (2) of Algorithm 3, however, we first need to find $\boldsymbol{P}_{\boldsymbol{k}}$.

Now, we use the Kronecker product property $\boldsymbol{a^T W b} = (\boldsymbol{b^T} \otimes \boldsymbol{a^T}) vec(\boldsymbol{W})$ to rewrite (44) as

$$[\hat{\boldsymbol{x}}(t+T) - \hat{\boldsymbol{x}}(t)]^T \hat{\boldsymbol{P}}_{\boldsymbol{k}} - 2\gamma^2 (\int_t^{t+T} \boldsymbol{x^T} \otimes \boldsymbol{d^T} d\tau) vec(\boldsymbol{G_{k+1}})$$
$$- 2(\int_t^{t+T} \boldsymbol{x^T} \otimes [(\boldsymbol{u} + \boldsymbol{F_k x})^T \boldsymbol{R}] d\tau) vec(\boldsymbol{F_{k+1}}) = \int_t^{t+T} \boldsymbol{x^T} \bar{\boldsymbol{Q}} \boldsymbol{x} d\tau \quad (48)$$

where the vectors $\hat{\boldsymbol{x}} \in \mathbb{R}^{\frac{n(n-1)}{2}}$ and $\hat{\boldsymbol{P}}_{\boldsymbol{k}} \in \mathbb{R}^{\frac{n(n-1)}{2}}$ are defined in the following forms

$$\hat{\boldsymbol{x}} = [x_1^2, 2x_1x_2, \cdots, 2x_1x_n, x_2^2, \cdots, 2x_{n-1}x_n, x_n^2]^T$$
$$\hat{\boldsymbol{P}}_{\boldsymbol{k}} = [P_{k(11)}, \cdots, P_{k(1n)}, P_{k(22)}, \cdots, P_{k(2n)}, \cdots, P_{k(nn)}]^T. \quad (49)$$

To solve $\boldsymbol{P_k}$, $\boldsymbol{G_{k+1}}$ and $\boldsymbol{F_{k+1}}$ in (48), define

$$\boldsymbol{d_x} = [\hat{\boldsymbol{x}}(t+T) - \hat{\boldsymbol{x}}(t), \cdots, \quad (50)$$
$$\hat{\boldsymbol{x}}(t+s_1T) - \hat{\boldsymbol{x}}(t+(s_1-1)T)]^T \in \mathbb{R}^{s_1 \times \frac{n(n-1)}{2}};$$
$$\boldsymbol{I_{xd}} = [\int_t^{t+T} (\boldsymbol{x} \otimes \boldsymbol{d}) d\tau, \cdots, \quad (51)$$
$$\int_{t+(s_1-1)T}^{t+s_1T} (\boldsymbol{x} \otimes \boldsymbol{d}) d\tau]^T \in \mathbb{R}^{s_1 \times np};$$
$$\boldsymbol{I_{xu}} = [\int_t^{t+T} \boldsymbol{x} \otimes (\boldsymbol{u} + \boldsymbol{F_k x}) d\tau, \cdots, \quad (52)$$
$$\int_{t+(s_1-1)T}^{t+s_1T} \boldsymbol{x} \otimes (\boldsymbol{u} + \boldsymbol{F_k x}) d\tau]^T \in \mathbb{R}^{s_1 \times nm};$$
$$\boldsymbol{\Phi} = [\boldsymbol{d_x}, -\boldsymbol{2I_{xd}}, -\boldsymbol{2I_{xu}}(\boldsymbol{I_n} \otimes \boldsymbol{R})]; \quad (53)$$
$$\boldsymbol{\Psi} = -[\int_t^{t+T} \boldsymbol{x^T} \bar{\boldsymbol{Q}} \boldsymbol{x} d\tau, \cdots \int_{t+(s_1-1)T}^{t+s_1T} \boldsymbol{x^T} \bar{\boldsymbol{Q}} \boldsymbol{x} d\tau]^T, \quad (54)$$

where integer $s_1 > 0$ is the sampling data group number. Then, the solution can be obtained by

$$\begin{bmatrix} \hat{\boldsymbol{P}}_{\boldsymbol{k}} \\ \text{vec}(\boldsymbol{G_{k+1}}) \\ \text{vec}(\boldsymbol{F_{k+1}}) \end{bmatrix} = (\boldsymbol{\Phi^T \Phi})^{-1} \boldsymbol{\Phi^T \Psi}. \quad (55)$$

Remark 9. To ensure that (55) achieves the unique solution, the persistence of the excitation condition needs to be satisfied. To this end, probing noise should be injected to control input u in (37). This is also called as exploration noise that does not affect the convergence [25]. In addition, the data group number s_1 should be no less than $\frac{n(n+1)}{2} + np + nm$, which is the number of unknown parameters to be calculated by (55).

In the next section, the correct performance of the proposed methods is validated by applying them to the lateral control of linearized F-16 dynamics.

5. Simulation Results

In this section, an example is given to verify the correct performance of proposed algorithms that solves optimal H_∞ static OPFB regulator problem. We used the F-16 linearized lateral dynamics at a particular flight condition given in example 5.3-1 of the book [41]. Parameters of the linearized F-16 lateral dynamics and the corresponding system vectors are

$$\begin{aligned}
\boldsymbol{A} &= \begin{bmatrix} -0.32 & 0.064 & 0.036 & -0.992 & 0 & 0.001 \\ 0 & 0 & 1 & 0.004 & 0 & 0 \\ -30.65 & 0 & -3.68 & 0.665 & -0.733 & 0.132 \\ 8.54 & 0 & -0.025 & -0.476 & -0.032 & -0.062 \\ 0 & 0 & 0 & 0 & -20.2 & 0 \\ 0 & 0 & 0 & 0 & 0 & -20.2 \end{bmatrix}; \\
\boldsymbol{B} &= \begin{bmatrix} 0 & 0 & 0 & 0 & 20.2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 20.2 \end{bmatrix}^T; \\
\boldsymbol{C} &= \begin{bmatrix} 0 & 0 & 0 & 57.2958 & 0 & 0 \\ 0 & 0 & 57.2958 & 0 & 0 & 0 \\ 57.2958 & 0 & 0 & 0 & 0 & 0 \\ 0 & 57.2958 & 0 & 0 & 0 & 0 \end{bmatrix}; \\
\boldsymbol{x} &= \begin{bmatrix} \beta & \phi & p & r & \delta_a & \delta_r \end{bmatrix}^T; \ \boldsymbol{u} = \begin{bmatrix} u_a & u_r \end{bmatrix}^T; \\
\boldsymbol{y} &= \begin{bmatrix} r & p & \beta & \phi \end{bmatrix}^T
\end{aligned} \tag{56}$$

where β denotes the side-slip, ϕ is the bank angle, p is the roll rate, r is the yaw rate, δ_a is the aileron actuator angle, δ_r is the rudder actuator angle, u_a is the aileron servo input, u_r is the rudder servo input. The factor of 57.2958 in the output-matrix $\boldsymbol{C}$ converts radians to degrees.

In a real-life scenario, the system states, and disturbances can be measured trough the sensors placed on the vehicle of interest. To exemplify, Inertial Measurement Units (IMUs) can be used to determine attitude (pitch-roll-yaw angles and their rates) of the vehicle, and structural health monitoring systems can be used to measure vibrational disturbances. In addition, the linearized model of F-16 (56) is only valid when the true speed is 502 $feet/sec$ and 300 psf dynamic pressure. When the aircraft changes its flight condition, i.e., the true speed or dynamics pressure values change, the Algorithm 3 can be used to obtain stabilizing Nash control policy without knowing the new system parameters of the F-16 assuming the system states, $\boldsymbol{x}$, is available. If there is a sensor problem and all of the system states are not available, then the Algorithm 4 can be used since it only requires the system output data $\boldsymbol{y}$.

Additionally, the objective functional (5) parameters are selected as $\boldsymbol{Q} = diag([50\ 100\ 100\ 50\ 0\ 0])$, $\boldsymbol{R} = \rho \times diag([0.1\ 0.7])$ with $\rho = 0.3$ for computation of the OPFB gain. Also, select the disturbance matrix as

$$\boldsymbol{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}^T, \tag{57}$$

and set the attenuation level $\gamma = 2.5$. To examine the robustness, assume that the system (1) experiences the worst-case disturbance (20). Realize that all of the Assumptions 1-3, and the sufficient condition $\boldsymbol{B R^{-1} B^T} \geq 2\gamma^{-2} \boldsymbol{D D^T}$ are satisfied with the given parameters in (56) and (57). Now, we first illustrate the performance of model-based Nash gain solutions (30) and (31), and then check whether the the Algorithms 1 and 2 converges the same game solutions as (30) and (31). After verifying the correctness of them, we illustrate the performance of the Nash solutions obtained in the Algorithms 3 and 4 that are model-free.

The Fig. 1 illustrates the zero control-effort response of the system (1). The Fig. 2 illustrates the optimal stabilizing gain (30) performance, which is derived in Theorem 3. The Nash OPFB gains found by (30) and (31) are

$$\boldsymbol{K_o^a} = \begin{bmatrix} -0.1395 & -0.8714 & 1.4785 & -1.0000 \\ -0.1410 & 0.0273 & -0.1070 & 0.0330 \end{bmatrix}, \tag{58}$$

$$\boldsymbol{N^o} = \begin{bmatrix} -0.0001 & -0.0006 & 0.0011 & -0.0007 \\ -0.0005 & 0.0001 & -0.0004 & 0.0001 \end{bmatrix}. \tag{59}$$

Now, to examine the performance of the Algorithms 1 and 2, we set the

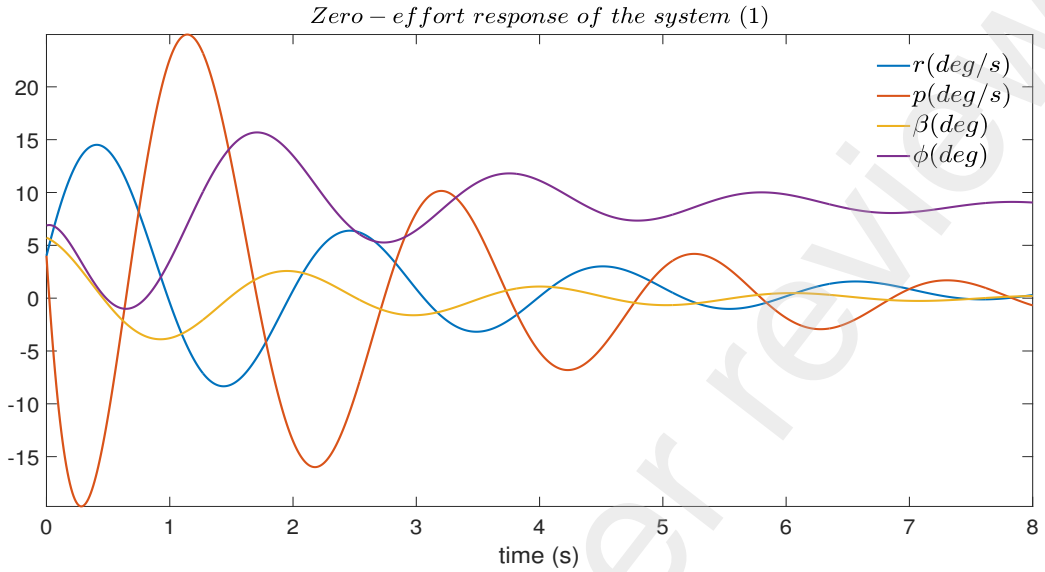

Figure 1: System response when no control is applied.

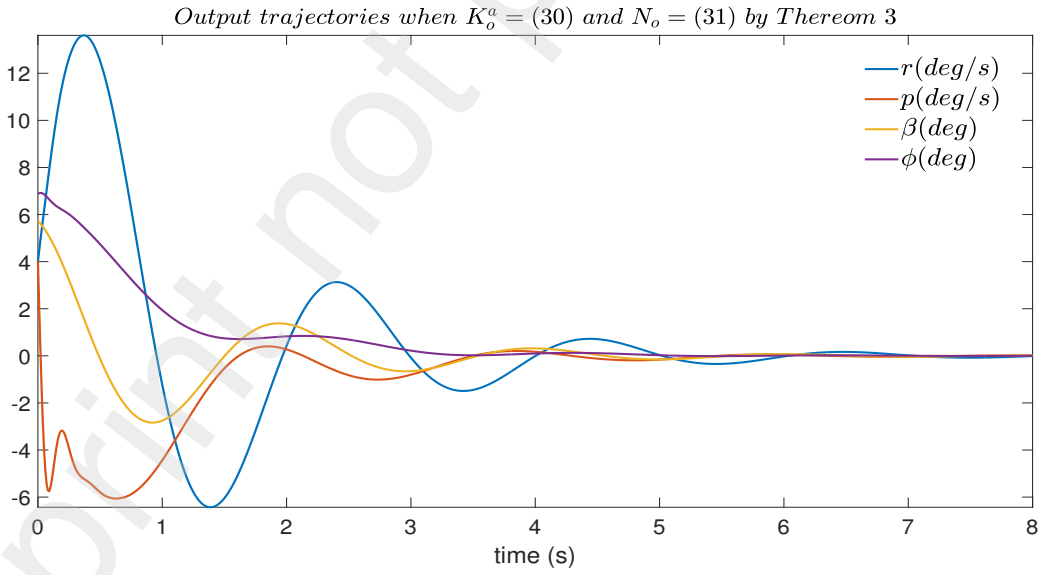

Figure 2: System response when the Nash gains (30) and (31) are employed.

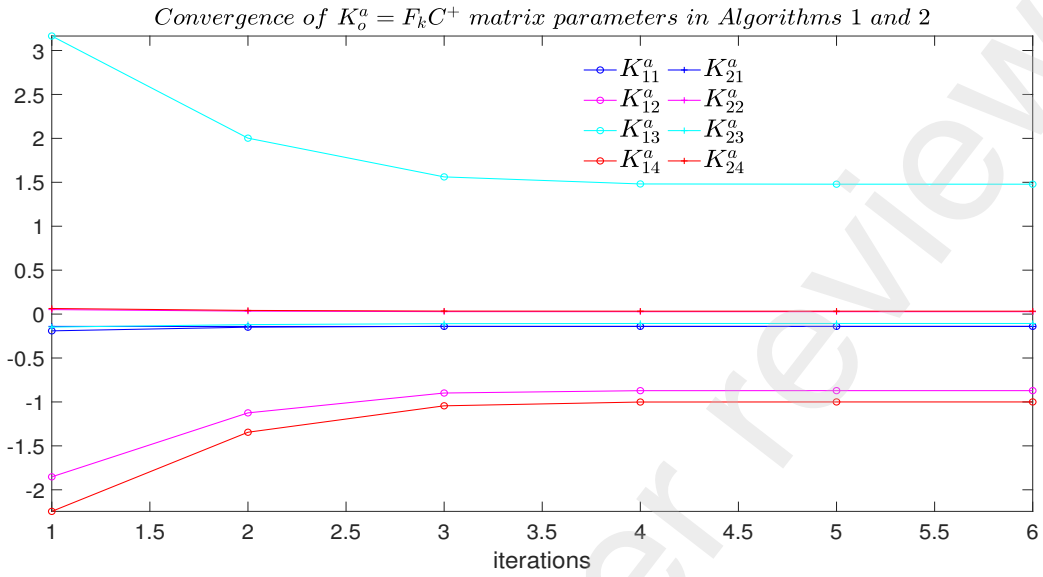

Figure 3: Convergence of the gain matrix $\boldsymbol{K_o^a}$ parameters by using the Algorithms 1 and 2.

initial sub-optimal stabilizing gain as

$$\boldsymbol{F_0} = \begin{bmatrix} -0.0888 & -0.1875 & 0.7076 & -0.2328 \\ -0.1382 & 0.0105 & -0.0884 & 0.0141 \end{bmatrix} C. \tag{60}$$

The convergence of optimal gain matrix parameters $\boldsymbol{K_o^a}$ can be observed from Fig. 3. Note that their convergence properties are exactly the same as each other and they converge to the same gain matrix (58) as expected. Therefore, the output trajectories figure with this resultant optimal stabilizing gain K^a is the same as Fig. 2. Now compare the system responses in Fig. 1 and Fig. 2 to observe the regulation performance. The convergence of disturbance gain matrix parameters $\boldsymbol{N_o}$ is also illustrated in Fig. 4. Note that the converged parameters of $\boldsymbol{N_o}$ are the same as the OPFB disturbance gain given in (59) as illustrated.

Next, to show the correct performance of the model-free Algorithm 3, we set the initial stabilizing gain F_0 as the same as (60). Then, based on the state information gathered from the system, the Algorithm 3 is employed to learn the Nash equilibrium gain solution $\boldsymbol{K_o^a}$ online. Once the IRL Algorithm converged, we applied the corresponding control $\boldsymbol{u} = -\boldsymbol{K_o^a y}$ using to the

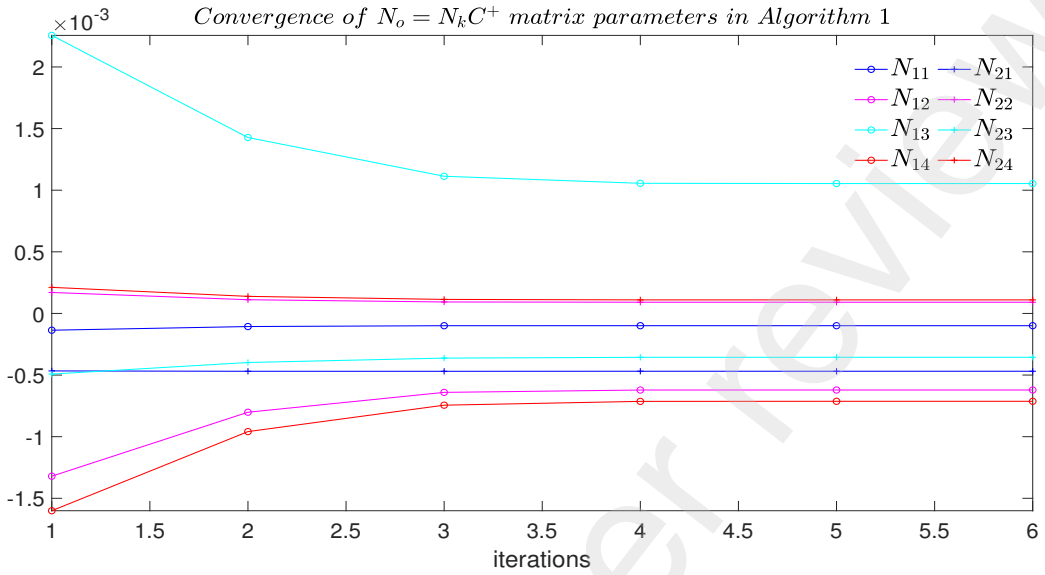

Figure 4: Convergence of the $\boldsymbol{N_o}$ matrix parameters by using the Algorithm 1.

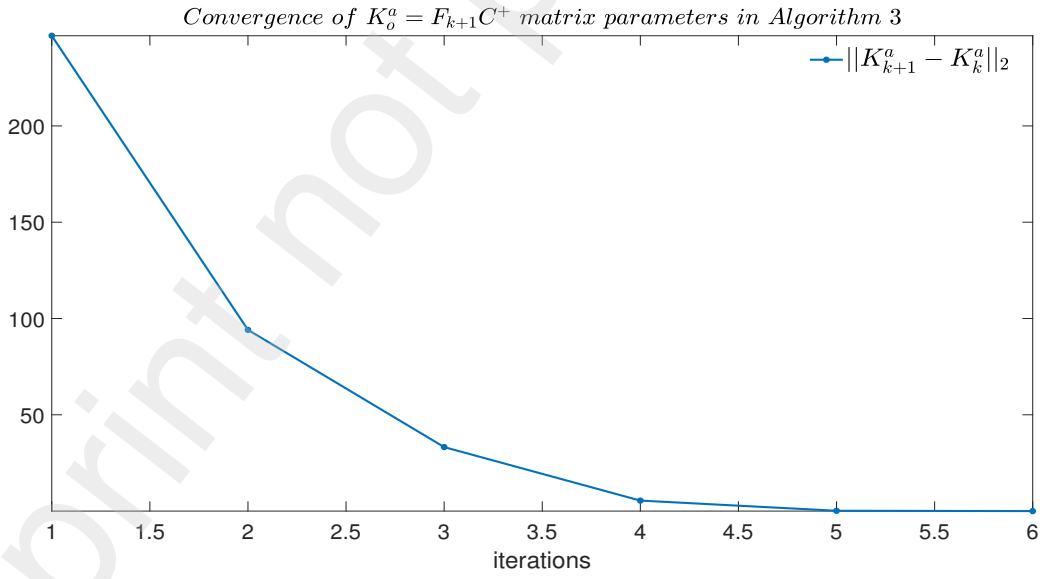

Figure 5: Convergence of the gain matrix $\boldsymbol{K_o^a}$ parameters by using the Algorithm 3.

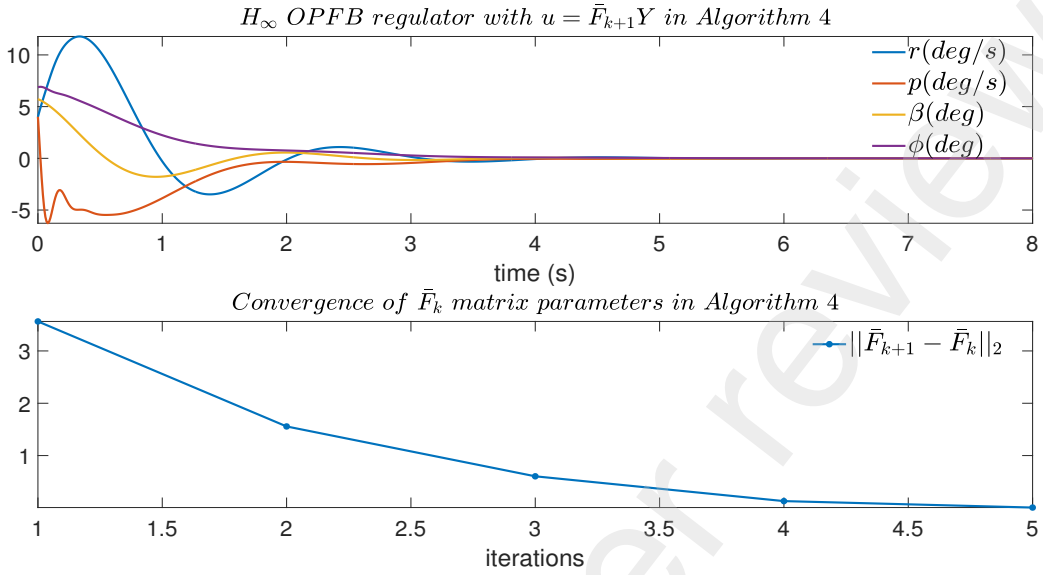

Figure 6: Performance of the Algorithm 4.

actual system (1). The Algorithm 3 is also converged to the same Nash gain $\boldsymbol{K_o^a}$ (58), thereby the resultant output vector states are the same as Fig. 2 as expected.

On the other hand, to check the correctness of the Algorithm 4, we select $\boldsymbol{Q} = k\boldsymbol{C^T C}$ with $k = 0.05$ as explained in Section 4.1. The resultant output trajectories are shown in Fig. 6. Additionally, we have calculated the Nash state feedback gain expression $\boldsymbol{K^*}$ given in (11), and also the observability matrix $\mathcal{O}$ for the system (1). Then, we compared the $\boldsymbol{K^*}$ with the $\bar{F}_k + 1\mathcal{O}$. It has been seen that the two matrices have almost the same elements and L_2 norm difference between them is calculated as 0.84, which verifies the correct performance of the Algorithm 4.

Lastly, the critical attenuation level obtained in the Theorems 2 and 3 is $\gamma^* = 0.05$. Note that with this critical attenuation γ^* level, the sufficient condition $\boldsymbol{BR^{-1}B^T} \geq 2(\gamma^*)^{-2}\boldsymbol{DD^T}$ is relaxed. However, this condition indeed required for the Kleinman based Algorithms 1-4, and the critical attenuation level is obtained as $\gamma^* = 0.49$, which is also compatible with the sufficient condition $\boldsymbol{BR^{-1}B^T} \geq 2(\gamma^*)^{-2}\boldsymbol{DD^T}$. Therefore, we conclude that the model-free algorithms reduce the L_2 gain performance.

6. Summary and Conclusions

In this paper, we proposed a novel augmented Hamiltonian, and develop new iterative algorithms based on stationarity conditions of the augmented Hamiltonian to obtain Nash equilibrium solutions. Additionally, the corresponding optimal gain solutions guarantee both stability and L_2 gain boundedness of an LTI system when the H_∞ static OPFB control method is employed. Convergence properties of two off-line iterative solution algorithms are given. Then, based on the Lyapunov iterations, an online off-policy IRL algorithm which is a model-free version of the offline Algorithm 2, is developed to solve the optimal H_∞ regulator problem by learning the Nash equilibrium solution (30) without requiring system state, control, and disturbance matrices. However, this Algorithm assumes the availability of the system state data. Thence, we come up with a novel model-free observer-based Algorithm 4 that only requires system output date to achieve stabilizing Nash control policies but it creates additional complexity as it requires more data to be collected to converge. Lastly, we applied the proposed algorithms to the linearized F-16 lateral dynamics at a particular flight condition to verify the correct performance of the proposed algorithms. Further research can be conducted to investigate how the state and output measurement delays affect the proposed methods.

7. Acknowledgements

This work is supported by the Office of Naval Research under Grant N00014-18-1-2221, the National Science Foundation under Grant ECCS-1839804 and Army Research Office under the Grant/Award Number W911NF-20-1-0132.

References

[1] J. Gadewadikar, F. L. Lewis, M. Abu-Khalaf, Necessary and sufficient conditions for h-infinity static output-feedback control, Journal of guidance, control, and dynamics 29 (4) (2006) 915–920.

[2] J. Gadewadikar, A. Bhilegaonkar, F. L. Lewis, Bounded l2 gain static output feedback: Controller design and implementation on an electromechanical system, IEEE Transactions On Industrial Electronics 54 (5) (2007) 2593–2599.

[3] A. Al-Tamimi, F. L. Lewis, M. Abu-Khalaf, Model-free q-learning designs for linear discrete-time zero-sum games with application to h-infinity control, Automatica 43 (3) (2007) 473–481.

[4] S. A. A. Rizvi, Z. Lin, Output feedback q-learning for discrete-time linear zero-sum games with application to the h-infinity control, Automatica 95 (2018) 213–221.

[5] Y. Jiang, K. Zhang, J. Wu, C. Zhang, W. Xue, T. Chai, F. L. Lewis, H-infinity based minimal energy adaptive control with preset convergence rate, IEEE Transactions on Cybernetics (2021).

[6] J. Han, H. Zhang, H. Jiang, X. Sun, H-infinity consensus for linear heterogeneous multi-agent systems with state and output feedback control, Neurocomputing 275 (2018) 2635–2644.

[7] J. Gadewadikar, F. L. Lewis, K. Subbarao, K. Peng, B. M. Chen, H-infinity static output-feedback control for rotorcraft, Journal of Intelligent and Robotic Systems 54 (4) (2009) 629–646.

[8] A. P. Valadbeigi, A. K. Sedigh, F. L. Lewis, H infinity static output-feedback control design for discrete-time systems using reinforcement learning, IEEE transactions on neural networks and learning systems 31 (2) (2019) 396–406.

[9] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[10] D. P. Bertsekas, Dynamic programming and optimal control: Vol. 1, Athena scientific Belmont, 2000.

[11] B. Kiumarsi, F. L. Lewis, H. Modares, A. Karimpour, M.-B. Naghibi-Sistani, Reinforcement q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics, Automatica 50 (4) (2014) 1167–1175.

[12] H. Modares, F. L. Lewis, Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning, IEEE Transactions on Automatic control 59 (11) (2014) 3051–3056.

[13] H. Modares, F. L. Lewis, Z.-P. Jiang, H-infinity tracking control of completely unknown continuous-time systems via off-policy reinforcement learning, IEEE transactions on neural networks and learning systems 26 (10) (2015) 2550–2562.

[14] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, F. L. Lewis, Optimal and autonomous control using reinforcement learning: A survey, IEEE transactions on neural networks and learning systems 29 (6) (2017) 2042–2062.

[15] C. Chen, H. Modares, K. Xie, F. L. Lewis, Y. Wan, S. Xie, Reinforcement learning-based adaptive optimal exponential tracking control of linear systems with unknown dynamics, IEEE Transactions on Automatic Control 64 (11) (2019) 4423–4438.

[16] B. Luo, H.-N. Wu, T. Huang, Off-policy reinforcement learning for h-infinity control design, IEEE transactions on cybernetics 45 (1) (2014) 65–76.

[17] M. Abu-Khalaf, F. L. Lewis, J. Huang, Neurodynamic programming and zero-sum games for constrained control systems, IEEE Transactions on Neural Networks 19 (7) (2008) 1243–1252.

[18] H. Zhang, Q. Wei, D. Liu, An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games, Automatica 47 (1) (2011) 207–214.

[19] Q. Wei, D. Liu, Q. Lin, R. Song, Adaptive dynamic programming for discrete-time zero-sum games, IEEE transactions on neural networks and learning systems 29 (4) (2017) 957–969.

[20] S. Mehraeen, T. Dierks, S. Jagannathan, M. L. Crow, Zero-sum two-player game theoretic formulation of affine nonlinear discrete-time systems using neural networks, IEEE transactions on cybernetics 43 (6) (2012) 1641–1655.

[21] L. M. Zhu, H. Modares, G. O. Peen, F. L. Lewis, B. Yue, Adaptive suboptimal output-feedback control for linear systems using integral reinforcement learning, IEEE Transactions on Control Systems Technology 23 (1) (2014) 264–273.

[22] F. A. Yaghmaie, S. Gunnarsson, F. L. Lewis, Output regulation of unknown linear systems using average cost reinforcement learning, Automatica 110 (2019) 108549.

[23] K. G. Vamvoudakis, F. L. Lewis, Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem, Automatica 46 (5) (2010) 878–888.

[24] H. Jiang, H. Zhang, X. Xie, Critic-only adaptive dynamic programming algorithms' applications to the secure control of cyber-physical systems, ISA transactions 104 (2020) 138–144.

[25] Y. Jiang, Z.-P. Jiang, Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics, Automatica 48 (10) (2012) 2699–2704.

[26] R. Moghadam, F. L. Lewis, Output-feedback h-infinity quadratic tracking control of linear systems using reinforcement learning, International Journal of Adaptive Control and Signal Processing 33 (2) (2019) 300–314.

[27] Q. Jiao, H. Modares, F. L. Lewis, S. Xu, L. Xie, Distributed l2-gain output-feedback control of homogeneous and heterogeneous systems, Automatica 71 (2016) 361–368.

[28] S. A. Arogeti, F. L. Lewis, Static output-feedback h_∞ control design procedures for continuous-time systems with different levels of model knowledge, IEEE Transactions on Cybernetics (2021).

[29] F. L. Lewis, D. Vrabie, V. L. Syrmos, Optimal control, John Wiley & Sons, 2012.

[30] J. Gadewadikar, F. L. Lewis, L. Xie, V. Kucera, M. Abu-Khalaf, Parameterization of all stabilizing h-infinity static state-feedback gains: application to output-feedback design, Automatica 43 (9) (2007) 1597–1604.

[31] W. M. Haddad, V. Chellaboina, Nonlinear dynamical systems and control, Princeton university press, 2011.

[32] B. M. Chen, Robust and H-infinity Control, Springer Science & Business Media, 2013.

[33] P. Apkarian, D. Noll, A. Rondepierre, Mixed h2 h-infinity control via nonsmooth optimization, SIAM Journal on Control and Optimization 47 (3) (2008) 1516–1546.

[34] T. Basar, P. Bernhard, H-infinity optimal control and related minimax design problems: a dynamic game approach, Springer Science & Business Media, 2008.

[35] D. Kleinman, On an iterative technique for riccati equation computations, IEEE Transactions on Automatic Control 13 (1) (1968) 114–115.

[36] H. W. Knobloch, A. Isidori, D. Flockerzi, Topics in control theory, Vol. 22, Birkhäuser, 2012.

[37] D. Moerder, A. Calise, Convergence of a numerical algorithm for calculating optimal output feedback gains, IEEE Transactions on Automatic Control 30 (9) (1985) 900–903.

[38] B. Datta, Numerical methods for linear control systems, Vol. 1, Academic Press, 2004.

[39] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, F. L. Lewis, Adaptive optimal control for continuous-time linear systems based on policy iteration, Automatica 45 (2) (2009) 477–484.

[40] C. Li, G. G. Yin, L. Guo, C.-Z. Xu, et al., State observability and observers of linear-time-invariant systems under irregular sampling and sensor limitations, IEEE Transactions on Automatic Control 56 (11) (2011) 2639–2654.

[41] B. L. Stevens, F. L. Lewis, E. N. Johnson, Aircraft control and simulation: dynamics, controls design, and autonomous systems, John Wiley & Sons, 2015.