2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

Extended-State Backward Iteration for Stackelberg Dynamic Games: Application to a 2-DOF Flexible Robot

Sami Elmadssia[1], Mohamad Saad[1], and Mourad Nedil[1]

Abstract— **This paper proposes a general framework for hierarchical dynamic games based on an *Iterative Derivation of Optimal Policies* (IDOP). The main theoretical result, stated in Theorem 1, reformulates the game using an *extended state* that includes the adjoint variables of all players. This enables a backward procedure in which the instantaneous optimal gain of each active player is computed while accounting for higher-priority strategies. A dedicated operator is introduced to compactly represent and solve the coupled Riccati equations arising from the Hamilton-Jacobi-Bellman framework. The method is generic and applicable to a broad class of hierarchical decision problems. Its effectiveness is demonstrated through two numerical examples and an experimental validation on a real two-degree-of-freedom (2-DOF) flexible serial robot.**

I. INTRODUCTION

Game theory provides a mathematical framework for analyzing decision-making in situations where multiple agents (players) interact, each seeking to optimize their own objective [1]–[3]. In contrast, optimal control theory focuses on determining control inputs that optimize a performance index subject to system dynamics [4]. When several controllers influence the same dynamical system, the problem becomes a dynamic game, blending optimal control principles with strategic interactions over time [5]. Such situations arise in numerous real-world applications [6], including homogeneous robot swarm [7], cyber-physical security scenarios involving attacker-defender dynamics [8], and human-robot collaboration [9]. In these contexts, agents may share the same state space but pursue different, and possibly conflicting, objectives, necessitating a generalization of classical single-agent optimal control into the framework of game-theoretic optimal control. The analysis and solution of such problems rely on advanced mathematical tools, notably Pontryagins Minimum Principle (PMP) for multiple players, which leads to coupled costate equations [10]; the Hamilton-Jacobi-Bellman (HJB) equations for dynamic games, which yield a set of coupled partial differential equations in the Nash framework [11]; and the theory of Linear Quadratic Differential Games (LQDG), in which solutions can be obtained via coupled Riccati differential equations [12]

In a Stackelberg game, the decision-making process is organized into two hierarchical roles: a leader and one or more followers. The leader selects its control action first, anticipating the optimal response of the follower(s), whereas each follower subsequently determines its own optimal strategy based on the leaders decision. This hierarchical interaction has been increasingly exploited in robotics. For instance, Stackelberg differential game formulations have been applied to modular robotic manipulators in human-robot collaboration [13], and to flexible two-arm surgical robots with advanced trajectory planning [14]. In the context of swarm robotic systems, such formulations often entail the solution of coupled HJB equations, arising from the interdependence of agents' strategies. Notably, adaptive robust Stackelberg control for uncertain mechanical systems has been developed in [15], while dynamic noncooperative game-theoretic approaches, including Stackelberg frameworks, have been explored in [16], [17], underscoring the growing significance of hierarchical game theory in the design and control of advanced robotic systems.

Inspired by these advances, we propose an *Iterative Derivation of Optimal Policies* (IDOP) from a Stackelberg-game perspective. A key aspect of our formulation is the construction of an *extended state* at each iteration, which augments the physical states of the system with the adjoint variables corresponding to the optimization problems of the other players. This augmented representation allows each player to anticipate the optimal responses of the others, and forms the basis of the backward iteration procedure described in Theorem 1.

At each step of this backward recursion, the control law for the active player is obtained while treating the strategies of higher-priority players as fixed, leading to a sequence of coupled Riccati equations. To facilitate the resolution of these coupled equations, we introduce the operator $\mathcal{G}$, which provides a compact reformulation and simplifies the iterative computation of optimal gains.

By combining the extended state framework, the operator $\mathcal{G}$, and the backward iteration mechanism, the proposed IDOP method yields closed-loop state-feedback gains that explicitly capture the leader-follower hierarchy and the dynamic coupling between the subsystems. The effectiveness of the proposed approach is demonstrated through two numerical examples and an experimental validation on a real two-degree-of-freedom (2-DOF) flexible robotic manipulator.

In contrast with classical Stackelberg and standard LQDG approaches, the novelty of this work lies in how the hierarchy is embedded into the optimal control formulation. The proposed IDOP method builds an extended state that includes the adjoint variables of higher-priority players, allowing each agent to anticipate their responses within the same framework. Combined with the backward iteration mechanism and the operator $\mathcal{G}$, this leads to a simplified sequence of

[1] Université du Québec en Abitibi-Témiscamingue, 445 Bd de l'Université, Rouyn-Noranda, QC J9X 5E4, Canada. sami.elmadssia@uqat.ca, mohamad.saad@uqat.ca, mourad.nedil@uqat.ca

978-1-6654-5784-2/26/$31.00 ©2026 IEEE

Riccati equations that explicitly encodes the leader–follower structure. This provides a compact and tractable alternative to conventional Stackelberg and LQDG formulations.

The rest of this paper is organized as follows. Section II presents the system model and defines the performance indices for the two robotic arms. Section III proposes the Stackelberg-based control strategies, followed by theoretical stability analysis. A simulation example is given in Section IV, and the conclusions are given in Section V.

II. Preliminaries and Problem Statement

A. Notations and Definition

The system dynamics are described by:

$$\dot{x}(t) = A^{(0)}x(t) + \sum_{\rho=1}^{m} B_\rho^{(0)}\, u_\rho(t), \quad x(0) = x_0, \quad t \geq 0, \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state vector available for feedback, and $u_\rho \in \mathbb{R}^{q_\rho}$ represents the control input of player ρ. The matrices $A^{(0)}$ and $B_\rho^{(0)}$ for $\rho = 1, \ldots, m$ denote the system and input matrices, respectively, with appropriate dimensions.

The players are organized in a hierarchical structure: u_1 corresponds to the leaders control, while u_ρ, for $\rho = 2, \ldots, m$, correspond to the followers controls. For notational convenience, $u_{-\rho}$ denotes the collection of control inputs of all players except player ρ, and $u_{+\rho}$ denotes the inputs of players whose indices are greater than ρ, that is, players $\rho+1, \rho+2, \ldots, m$.

Definition 1. [18], [19] An hierarchical Stackelberg game is a dynamic game with m agents arranged in a strict priority order. Each agent $j \in \{1, \ldots, m\}$ aims to minimize its own cost functional $J_j(u_j, u_{-j})$ subject to its dynamics.
The equilibrium is computed in a leader-follower manner, where higher-priority agents anticipate the optimal responses of lower-priority agents.

Definition 2. [18], [19] The optimal control policy $u_i^*(t)$ for agent i is the strategy that minimizes its cost functional J_i, considering the hierarchical structure.

This work presents an innovative approach to compute the optimal feedback gain matrix in hierarchical control settings. By leveraging an iterative scheme, the method eliminates the need to solve complex coupled Riccati equations typically encountered in Stackelberg differential games.

B. General Framework

The proposed approach is based on decomposing the hierarchical Stackelberg game into a sequence of nested subgames, each corresponding to a player u_j in a predetermined priority hierarchy.

We consider, for each player j, the following performance index:

$$\begin{aligned} J_j(u_j) &= \frac{1}{2} x^{(0)T}(t_f)\, \bar{P}_j\, x^{(0)}(t_f) \\ &+ \frac{1}{2}\int_t^{t_f} L_j^{(0)}\big(x^{(0)}(\tau), u_j(\tau), u_{-j}(\tau)\big)\, d\tau, \end{aligned}$$

where the state trajectory starts from $x^{(0)}(t) = x(t)$, the terminal time $t_f > 0$ is fixed, and the weighting matrix $\bar{P}_j \in \mathbb{R}^{n\times n}$ satisfies $\bar{P}_j \succ 0$. The stage cost function is given by:

$$\begin{aligned} L_j^{(0)}(.) &= x^{(0)T}(t) Q_j^{(0)} x^{(0)}(t) \\ &+ u_j^T(t) R_{j,j}\, u_j + \sum_{\rho=1}^{m} u_\rho^T(t) R_{j,\rho}\, u_\rho(t), \end{aligned}$$

where $Q_j^{(0)} \succeq 0$ and $R_{j,\rho} \succeq 0$ for all $j, \rho = 1, \ldots, m$.

Let $u_1 \in \mathcal{U}_1$ denote the control policy of the leader, and let $u_j \in \mathcal{U}_j$ denote the control policy of each follower $j \in \{2, \ldots, m\}$. Each player $j \in \{1, \ldots, m\}$ aims to minimize their individual cost functional $J_j(u_1, u_2, \ldots, u_m)$. Given a strategy u_1 announced by the leader, each follower $j \in \{2, \ldots, m\}$ selects an optimal strategy u_j^* that minimizes their individual cost, under the assumption that all other followers also act optimally. The corresponding optimality condition is expressed as follows:

$$J_j(u_1, u_j^*, u_{-j}^*) \leq J_j(u_1, u_j, u_{-j}^*), \quad \forall u_j \in \mathcal{U}_j. \tag{2}$$

Definition 3. [18], [19] A strategy profile $(u_1^*, u_2^*, \ldots, u_m^*)$ is said to be a *Stackelberg equilibrium* if the following conditions hold:

$$J_j(u_j^*, u_{-j}^*) \leq J_j(u_j, u_{-j}^*), \forall u_j \in \mathcal{U}_j. \tag{3}$$

The leader anticipates these optimal responses and selects a strategy u_1^* that minimizes its own cost function:

$$u_1^* \in \arg\min_{u_1 \in \mathcal{U}_1} J_1(u_1, u_2^*(u_1), \ldots, u_m^*(u_1)), \tag{4}$$

where each follower's best response $u_j^*(u_1)$ is defined by:

$$u_j^*(u_1) \in \arg\min_{u_j \in \mathcal{U}_j} J_j(u_1, u_2^*(u_1), \ldots, u_j, \ldots, u_m^*(u_1)). \tag{5}$$

In the continuous-time open-loop information setting considered in this work, the players in the Stackelberg game select their entire control trajectories $u_j(t)$, $t \in [0, T]$, at the initial time $t = 0$ and keep them fixed over the whole horizon. Under this assumption, our objective is to determine

$$J_j^* = \min_{u_j(\cdot)} J_j\big(u_j(\cdot), u_{-j}(\cdot)\big). \tag{6}$$

III. Iterative Derivation of Optimal Policies

A. Core Idea

The core idea underlying the proposed hierarchical strategy design is to recursively construct an augmented system state that encapsulates the adjoint variables associated with each player's optimization problem. This recursive augmentation enables each follower to anticipate the actions of higher-priority players and adapt its strategy accordingly, while the leader optimizes its policy by accounting for the followers' best responses. The lifting operator formalism provides a rigorous and systematic framework for structuring this hierarchical interdependence and for deriving optimal state-feedback control laws. To operationalize this framework, we introduce a novel iterative method, termed the IDOP, which solves the optimal control problem for each

player u_i in the form of a state-feedback policy. At each iteration i, the method computes the optimal control law u_i^* by solving a decoupled optimization problem. By replacing the optimal control $u_i^*(t)$ into the state equations at each iteration, the updated state-space form at iteration i becomes:

$$\dot{x}^{(i)}(t) = A^{(i)}x^{(i)}(t) + \sum_{\varrho=i+1}^{m} B_{\varrho}^{(i)}\, u_{\varrho}(t), \tag{7}$$

where

$$x^{(i)^T}(t) = \left[\begin{array}{cc} x^{(i-1)^T}(t) & \lambda_i^T(t) \end{array}\right]^T, \tag{8}$$

$$A^{(i)} = \left[\begin{array}{cc} A^{(i-1)} & -B_i^{(i-1)}R_{i,i}^{-1}B_i^{(i-1)^T} \\ -Q_i^{(i-1)} & -A^{(i-1)^T} \end{array}\right] \tag{9}$$

$$B_{\varrho}^{(i)} = \left[\begin{array}{c} B_{\varrho}^{(i-1)} \\ 0 \end{array}\right],\ i = 1, \ldots, \varrho. \tag{10}$$

This recursive construction successively augments the system state and dynamics at each level, reflecting the hierarchical structure of the decision process. Specifically, we define the Hamiltonian $\mathcal{H}_i^{(i-1)}$ associated with follower i, $i = 1, \ldots, m$, as:

$$\begin{aligned} \mathcal{H}_i^{(i-1)}(.) &= \frac{1}{2}L_i^{(i-1)}(.) + \lambda_i^T\dot{x}^{(i)}(t) \\ &= \frac{1}{2}L_i^{(i-1)}(.) + \lambda_i^T\Big(A^{(i-1)}x^{(i-1)} \\ &+ B_i^{(i-1)}u_i + \sum_{\varrho=i+1}^{m} B_{\varrho}^{(i-1)}u_{\varrho}(t)\Big) \end{aligned}$$

where

$$\begin{aligned} L_j^{(i-1)}(.) &= x^{(i-1)^T}(t)Q_j^{(i-1)}x^{(i-1)}(t) \\ &+ u_j^T(t)R_{j,j}u_j(t) + \sum_{\varrho=j+1}^{m} u_{\varrho}^T(t)R_{j,\varrho}u_{\varrho}(t). \end{aligned}$$

with

$$Q_j^{(i)} = \left[\begin{array}{cc} Q_j^{(i-1)} & 0 \\ 0 & B_i^{(i-1)}R_{i,i}^{-1}R_{j,i}R_{i,i}^{-1}B_i^{(i-1)^T} \end{array}\right]. \tag{11}$$

and $\lambda_i(t)$ denotes the costate vector associated with the state $x^{(i)}(t)$.

The partial derivative of $\mathcal{H}_i^{(i-1)}$ with respect to u_i yields:

$$\begin{aligned} \frac{\partial\mathcal{H}_i^{(i-1)}}{\partial u_i} &= \frac{1}{2}\frac{\partial}{\partial u_i}L_i\left(x^{(i-1)}, u_i, u_{+i}\right) \\ &+ \lambda_i^T\frac{\partial}{\partial u_i}\Big(A^{(i-1)}x^{(i-1)} + B_i^{(i-1)}u_i \\ &+ \sum_{\varrho=i+1}^{m} B_{\varrho}^{(i-1)}u_{\varrho}(t)\Big) \\ &= R_{i,i}u_i(t) + B_i^{(i-1)T}\lambda_i(t). \end{aligned}$$

The optimality condition $\frac{\partial\mathcal{H}_i^{(i-1)}}{\partial u_i} = 0$ leads to the following expression for the optimal control law:

$$u_i^*(t) = -R_{i,i}^{-1}B_i^{(i-1)T}\lambda_i(t), \quad i = 1, \ldots, m-1. \tag{12}$$

The dynamics of the costate variables are obtained from:

$$\dot{\lambda}_i(t) = -\left(\frac{\partial\mathcal{H}_i^{(i-1)}}{\partial x^{(i)}}\right)^T = -Q_i^{(i-1)}x^{(i-1)}(t) - A^{(i-1)T}\lambda_i(t), \tag{13}$$

$i = 1, \ldots, m-1$. It lays the foundation for deriving the optimal feedback policies in the next section.

B. Recursive Lifting Framework and Optimal Feedback Computation

We introduce $\mathcal{T}^{(i-j)}(t)$ as the *recursive lifting operator*, which transforms the original state $x^{(i)}(t)$ into an augmented state $x^{(j)}(t)$ through a sequence of successive augmentations. The initial case is defined by $\mathcal{T}^{(0)}(t) \triangleq I_n$, and for $i = 1, \ldots, m$ we have:

$$\mathcal{T}^{(i-j)}(t) \triangleq \frac{\partial x^{(i)}(t)}{\partial x^{(i-1)}(t)} \otimes \frac{\partial x^{(i-1)}(t)}{\partial x^{(i-2)}(t)} \otimes \cdots \frac{\partial x^{(j+1)}(t)}{\partial x^{(j)}(t)}, \tag{14}$$

where $\otimes$ is the Kronecker product.

This recursive definition forms the basis for constructing the dynamics of the extended system and for obtaining the optimal control strategies in an iterative manner. The main theoretical contribution presented here establishes how these recursive relations enable the computation of optimal feedback laws for hierarchical Stackelberg differential games.

Remark It is worth noting that the lifting operator defined in (14) is *independent* of the current extended state $x^{(i)}(t)$, $i = 0, \ldots, m$. Indeed, $\mathcal{T}^{(i)}(t)$ depends solely on the matrices $P_1(t), P_2(t), \ldots, P_i(t)$ obtained from the corresponding Riccati equations, and not on the instantaneous values of the state variables.

Theorem 1. Consider the system dynamics described in equation (1). For each player $i \in \{1, \ldots, m\}$, the optimal state-feedback control law takes the form:

$$u_i^*(t) = -R_{i,i}^{-1}\, B_i^{(i-1)T}\, \mathcal{G}^{(i-1)}(t)\, x^{(i-1)}(t). \tag{15}$$

where $\mathcal{G}^{(i-1)}(t) = P_i(t)\,\mathcal{T}^{(i-1)}(t)$ denotes the composite gain matrix, which satisfies the following differential Riccati equation:

$$\begin{aligned} \dot{\mathcal{G}}^{(i-1)}(t) &= -A^{(i-1)T}\mathcal{G}^{(i-1)}(t) - \mathcal{G}^{(i-1)}(t)A^{(i-1)} \\ &+ \mathcal{G}^{(i-1)}(t)\sum_{\varrho=i}^{m} B_{\varrho}^{(i-1)}R_{\varrho,\varrho}^{-1}B_{\varrho}^{(\varrho-1)T} \\ &\times \mathcal{G}^{(\varrho-1)}(t)\,\mathcal{T}^{(\varrho-i)}(t) - Q_i^{(i-1)}. \end{aligned} \tag{16}$$

Proof . Let us represent the adjoint variable $\lambda_i(t)$ as a time-dependent linear mapping of the extended state:

$$\lambda_i(t) = \mathcal{G}^{(i-1)}(t)\, x^{(i-1)}(t), \tag{17}$$

where $\mathcal{G}^{(i-1)}(t)$ is the instantaneous gain associated with the state $x^{(i-1)}(t)$. For $i = 1, \ldots, m$, differentiating $\lambda_i(t)$ with respect to time yields:

$$\dot{\lambda}_i(t) = \dot{\mathcal{G}}^{(i-1)}(t)\, x^{(i-1)}(t) + \mathcal{G}^{(i-1)}(t)\,\dot{x}^{(i-1)}(t), \tag{18}$$

$$-Q_i^{(i-1)}x^{(i-1)}(t) - \left(A^{(i-1)}\right)^T \lambda_i(t) = \dot{\mathcal{G}}^{(i-1)}(t)\,x^{(i-1)}(t) + \mathcal{G}^{(i-1)}(t)\left(A^{(i-1)}x^{(i-1)}(t) + \sum_{\varrho=i}^{m} B_\varrho^{(i-1)}\,u_\varrho(t)\right) \tag{19}$$

using (13) and (7) we get (19)

By substituting the expression of $\lambda_i(t)$ from (17) into (19), and replacing $u_\varrho(t)$ with its optimal form $u_\varrho^*(t)$, given by

$$\begin{aligned} u_\varrho^*(t) &= -R_{\varrho,\varrho}^{-1}\,B_\varrho^{(\varrho-1)T}\,\mathcal{G}^{(\varrho-1)}(t)\,x^{(\varrho-1)}(t). \\ &= -R_{\varrho,\varrho}^{-1}\,B_\varrho^{(\varrho-1)T}\,\mathcal{G}^{(\varrho-1)}(t)\mathcal{T}^{(\varrho-i)}(t)x^{(i-1)}(t). \end{aligned}$$

we recover the expression in (16).

IV. Application Examples

A. Numerical Example

To demonstrate the behavior of the coupled Riccati equations for a two-player dynamic game, we examine the following linear time-invariant (LTI) system:

$$\dot{x}(t) = A^{(0)}x(t) + B_1^{(0)}u_1(t) + B_2^{(0)}u_2(t), \tag{20}$$

with

$$A^{(0)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -2 & -0.5 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & -2 & -0.5 \end{bmatrix}, B_1^{(0)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, B_2^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

The evolution of the diagonal and off-diagonal entries of $P_1(t)$ and $P_2(t)$ is illustrated in Figures 1–4. The results indicate that the trajectories converge smoothly to their steady-state values, highlighting the stabilizing properties of the feedback gains:

$$K_1(t) = R_{1,1}^{-1}B_1^{(0)T}\mathcal{G}^{(0)}(t), \quad K_2(t) = R_{2,2}^{-1}B_2^{(1)T}\mathcal{G}^{(1)}(t). \tag{21}$$

and the optimal feedback laws are:

$$\begin{aligned} u_1^*(t) &= -K_1(t)x(t) \\ u_2^*(t) &= -K_2(t)x^{(1)}(t) \end{aligned} \tag{22}$$

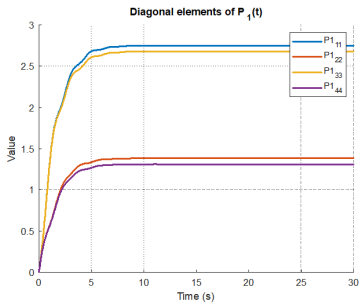

Fig. 1. Trajectory over time of the diagonal entries of $P_1(t)$.

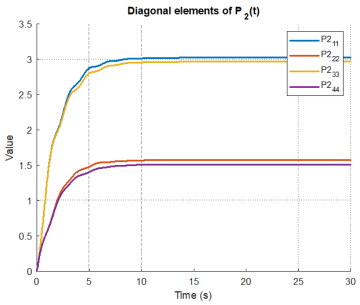

Fig. 2. Trajectory over time of the diagonal entries of $P_2(t)$.

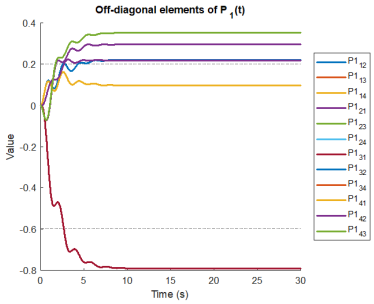

Fig. 3. Time evolution of the off-diagonal elements of $P_1(t)$.

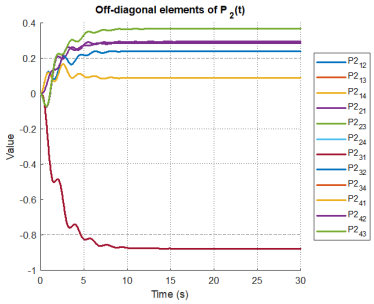

Fig. 4. Time evolution of the off-diagonal elements of $P_2(t)$.

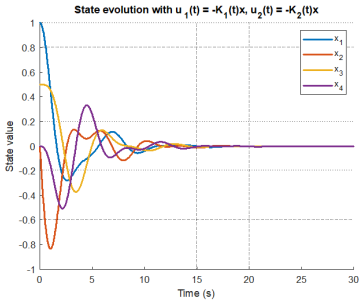

Fig. 5. Simulated state trajectories under the feedback laws .

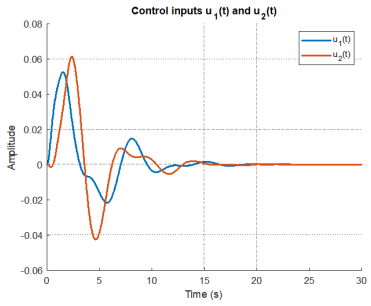

Fig. 6. Time evolution of the control inputs $u_1(t)$ and $u_2(t)$.

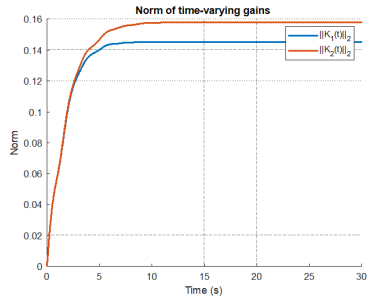

Fig. 7. Time evolution of the norm of time-varying gains.

B. Application to the Control of a 2-DOF Serial Flexible Robot

In this study, the experimental setup is a two-degree-of-freedom (2-DOF) serial flexible manipulator manufactured by Quanser [20] (see Fig. 8). The system includes two revolute joints driven by DC motors (identified as joints 1 and 2), which are connected through flexible-link (labeled as links 7 and 8), with elastic coupling at the joints. The setup includes optical encoders (3 and 4) for angular position sensing, amplifiers and power interfaces (5, 6, 12), and structural supports mounted on a base (14). The first joint (leader) drives the base link, while the second joint (follower) is responsible for the distal segment. This configuration is well suited to evaluate hierarchical control strategies under structural flexibility constraints. The input signal I_{m_1} represents the current supplied to the first motor. Accordingly, the dynamics of the Stage 1 subsystem of the 2-DOF serial flexible (2DSF) manipulator can be expressed as:

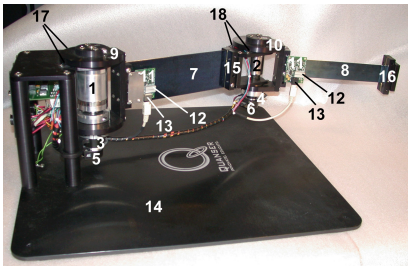

Fig. 8. Experimental 2-DOF flexible-link robotic arm.

$$\dot{x}_1 = A_1 x_1 + B_1 I_{m_1}, \tag{23}$$

TABLE I

COMPONENT NOMENCLATURE FOR THE EXPERIMENTAL 2-DOF FLEXIBLE-LINK ROBOTIC ARM

ID #	Description	ID #	Description
1	Harmonic Drive	2	Harmonic Drive
3	DC Motor	4	DC Motor
5	Motor #1 Encoder	6	Motor
7	Flexible Link	8	Flexible Link
9	Rigid Joint	10	Rigid Joint
11	Amplifier Board	12	Offset Potentiometer
13	Connector	14	Base Plate
15	Link End-Effector	16	Link #2 End-Effector
17	Joint Limit Switches	18	Joint #2 Limit Switches

where

$$x_1^T = \begin{bmatrix} \theta_{11}(t) & \theta_{12}(t) & \dot{\theta}_{11}(t) & \dot{\theta}_{12}(t) \end{bmatrix},$$

and the dynamics of the Stage 2 subsystem of the 2-DOF serial flexible (2DSF) manipulator can be expressed as:

$$\dot{x}_2 = A_2 x_2 + B_2 I_{m_2},$$

where I_{m_2} represents the current supplied to the second motor and

$$x_2^T = \begin{bmatrix} \theta_{21}(t) & \theta_{22}(t) & \dot{\theta}_{21}(t) & \dot{\theta}_{22}(t) \end{bmatrix}.$$

The overall dynamics can be expressed in the state-space form of equation (1) as

$$\dot{x}(t) = A^{(0)} x(t) + B_1^{(0)} I_{m_1}(t) + B_2^{(0)} I_{m_2}(t) \tag{24}$$

where

$$\begin{aligned} A^{(0)} &= \text{blkdiag}(A_1, A_2) \\ B_1^{(0)} &= \begin{bmatrix} B_1 \\ 0 \end{bmatrix} \text{ and } B_2^{(0)} = \begin{bmatrix} 0 \\ B_2 \end{bmatrix}. \end{aligned}$$

The system consists of two flexible stages represented by the following matrices:

Stage 1:

$$A_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 2018.4 & -91.7 & 0 \\ 0 & -2147.2 & -97.6 & 0 \end{bmatrix}, \quad B_{11} = \begin{bmatrix} 0 \\ 0 \\ 10.9 \\ -10.9 \end{bmatrix}.$$

Stage 2:

$$A_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 827.5 & 0 & 28.3 \\ 0 & -1215.4 & 0 & -13.3 \end{bmatrix}, \quad B_{22} = \begin{bmatrix} 0 \\ 0 \\ 7.75 \\ -7.75 \end{bmatrix}.$$

The leader-follower hierarchy is especially relevant in this system, where the control action of the second arm must adapt to and follow the motion induced by the first arm. As such, the first arm is viewed as the leader, while the second arm acts as a follower, leading to a natural Stackelberg formulation of the control problem. For simulation, we consider the initial condition:

$$x(0) = \begin{bmatrix} 5^\circ & -5^\circ & 0 & 0 & 5^\circ & 5^\circ & 0 & 0 \end{bmatrix}^T. \tag{25}$$

The coupled closed-loop dynamics:

$$\dot{x} = A_0 x + B_1^{(0)} u_1(t) + B_2^{(0)} u_2(t). \tag{26}$$

Figure 9 shows the joint angle evolution for each stage, while Fig. 10 shows the motor currents $I_{m1}(t)$ and $I_{m2}(t)$. Both joints converge to the equilibrium with control inputs respecting the leaderfollower hierarchy.

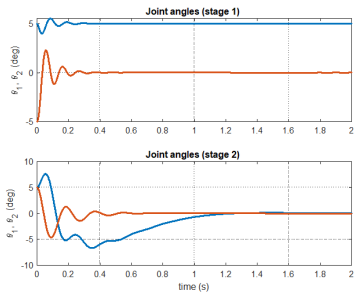

Fig. 9. Joint angles $\theta_1(t), \theta_2(t)$ for each stage under hierarchical DRE control.

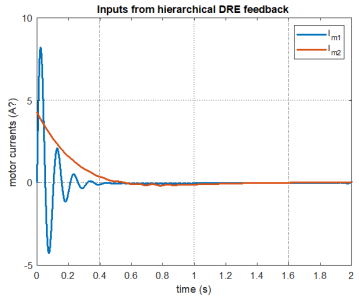

Fig. 10. Motor currents $I_{m1}(t)$ and $I_{m2}(t)$ generated by the hierarchical control laws.

V. Conclusion

In this paper, a hierarchical control strategy based on the Stackelberg differential game framework has been presented for multi-input dynamic systems as described in (1). By solving coupled Riccati equations derived from the Hamilton-Jacobi-Bellman formulation, closed-loop extended state-feedback gains were obtained, ensuring stability while explicitly accounting for asymmetric coupling effects between subsystems.

The proposed *Iterative Determination of Optimal Policies* (IDOP) algorithm efficiently computed optimal responses without requiring complete knowledge of the system model. Simulation studies demonstrated that the method achieves high-precision trajectory tracking and maintains robust performance under coupling conditions.

Future research will aim to extend the proposed approach to H_∞ formulations for enhanced disturbance rejection, incorporate nonlinear compensation to mitigate model uncertainties, and explore reinforcement learning-based strategies for real-time adaptation in multi-agent systems. Moreover, the extended state formulation will be employed to estimate the costate variables using the operators $\mathcal{T}^{(i)}$ and $\mathcal{G}^{(i)}$, thereby providing a systematic and effective framework for costate reconstruction in complex hierarchical control problems.

References

[1] Myerson, R. B., *Game Theory: Analysis of Conflict*. Harvard University Press, 1991.

[2] Barron, E. N., *Game Theory: An Introduction*, 3rd ed. Wiley, 2024.

[3] Marden, J. R., and Shamma, J. S., "Game Theory and Control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 105–134, 2018.

[4] Kirk, D. E., *Optimal Control Theory: An Introduction*. Dover Publications, 2004.

[5] H. Tembine, Q. Zhu, and T. Başar, "Risk-sensitive mean-field games," *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 835–850, Apr. 2014, doi: 10.1109/TAC.2013.2289711.

[6] Li, Z., and Wang, J., "Cooperative Control of Multi-Agent Systems with Applications in Unmanned Aerial Vehicles," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 4, pp. 3223–3233, 2018.

[7] X. Li, R. Zhou, G. Sun, Y. Zhang, and Z. Wang, "Distributed multiple shape formation in homogeneous robot swarms," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 18341–18359, 2025, doi: 10.1109/TASE.2025.3586130.

[8] H. Mejdi, S. Elmadssia, M. Koubaa and T. Ezzedine, "A Comprehensive Survey on Game Theory Applications in Cyber-Physical System Security: Attack Models, Security Analyses, and Machine Learning Classifications," in IEEE Access, vol. 12, pp. 163638-163653, 2024.

[9] Y. Cui, T. An, B. Dong, B. Ma, and Z. Zhang, "Bilayer nonzero-sum differential game-based optimal control of modular robot manipulator for human-robot collaboration," *European Journal of Control*, vol. 83, p. 101225, 2025, doi: 10.1016/j.ejcon.2025.101225.

[10] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, International Series of Monographs in Pure and Applied Mathematics. New York, NY, USA: Interscience, 1962.

[11] Yong, J., and Zhou, X. Y., *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer, 1999.

[12] Engwerda, J. C., *LQ Dynamic Optimization and Differential Games*. Wiley, 2005.

[13] T. An, X. Zhu, B. Ma, and H. Jiang, "Hierarchical approximate optimal interaction control of human-centered modular robot manipulator systems: A Stackelberg differential game-based approach," *Neurocomputing*, vol. 585, p. 127573, 2024.

[14] Y. Xie, X. Zhao, Y. Jiang, Y. Wu, and H. Yu, "Flexible control and trajectory planning of medical two-arm surgical robot," *Frontiers in Neurorobotics*, vol. 18, p. 1451055, 2024.

[15] Q. Sun, X. Wang, G. Yang, and Y. Liu, "Optimal parameter selection for constraint-following control for mechanical systems based on Stackelberg game," *Nonlinear Dynamics*, vol. 109, pp. 1629–1650, Aug. 2022, doi: 10.1007/s11071-022-07512-5.

[16] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, 2nd ed. Philadelphia, PA, USA: SIAM, 1999.

[17] H. Abou-Kandil and P. Bertrand, "Analytical solution for an open-loop Stackelberg game," *IEEE Trans. Autom. Control*, vol. 30, no. 12, pp. 1222–1224, Dec. 1985.

[18] A. Bensoussan, S. Chen, and S. P. Sethi, "The maximum principle for global solutions of stochastic Stackelberg differential games," *arXiv preprint arXiv:1210.3124*, 2012. [Online]. Available: `https://arxiv.org/abs/1210.3124`

[19] K. G. Vamvoudakis, D. Cansever, F. L. Lewis, M. A. Demetriou, and H. Modares, "Open-loop Stackelberg learning solution for hierarchical cyber-physical energy systems," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 12, pp. 1795–1814, 2019.

[20] Quanser Inc., *2-DOF Serial Flexible Link Robot Reference Manual*, Quanser Inc., Markham, ON, Canada, 2020.