SIAM J. OPTIM.
Vol. 34, No. 4, pp. 3723–3749

THE COMPUTATION OF APPROXIMATE FEEDBACK STACKELBERG EQUILIBRIA IN MULTIPLAYER NONLINEAR CONSTRAINED DYNAMIC GAMES*

JINGQI LI†, SOMAYEH SOJOUDI†, CLAIRE J. TOMLIN†, AND DAVID FRIDOVICH-KEIL‡

Abstract. Solving feedback Stackelberg games with nonlinear dynamics and coupled constraints, a common scenario in practice, presents significant challenges. This work introduces an efficient method for computing approximate local feedback Stackelberg equilibria in multiplayer general-sum dynamic games, with continuous state and action spaces. Different from existing (approximate) dynamic programming solutions that are primarily designed for unconstrained problems, our approach involves reformulating a feedback Stackelberg dynamic game into a sequence of nested optimization problems, enabling the derivation of Karush–Kuhn–Tucker (KKT) conditions and the establishment of a second-order sufficient condition for local feedback Stackelberg equilibria. We propose a Newton-style primal-dual interior point method for solving constrained linear quadratic (LQ) feedback Stackelberg games, offering provable convergence guarantees. Our method is further extended to compute local feedback Stackelberg equilibria for more general nonlinear games by iteratively approximating them using LQ games, ensuring that their KKT conditions are locally aligned with those of the original nonlinear games. We prove the exponential convergence of our algorithm in constrained nonlinear games. In a feedback Stackelberg game with nonlinear dynamics and (nonconvex) coupled costs and constraints, our experimental results reveal the algorithm's ability to handle infeasible initial conditions and achieve exponential convergence toward an approximate local feedback Stackelberg equilibrium.

Key words. feedback Stackelberg equilibrium, dynamic games, mathematical programming

MSC codes. 49K99, 68Q25, 91A25

DOI. 10.1137/24M1634709

1. Introduction.

Dynamic game theory [5] provides tools for analyzing strategic interactions in multiagent systems. It has broad applications in control [10], biology [26], and economics [17]. A well-known equilibrium concept in dynamic game theory is the *Nash equilibrium* [36], where players pursue strategies that are unilaterally optimal, and players make decisions simultaneously. However, this may not apply to a broad class of games where a decision hierarchy exists, such as lane-merging in highway driving [52], predator-prey competition in biology [2], and retail markets in economics [30]. These games could be more naturally formulated as *Stackelberg games* [47], where players act sequentially in a predefined order. For such games, the *Stackelberg equilibrium* is the appropriate equilibrium concept.

The formulation of Stackelberg equilibria depends on the information structure [5]. For instance, in scenarios where players lack access to the current game state, one

*Received by the editors January 29, 2024; accepted for publication (in revised form) September 3, 2024; published electronically December 2, 2024.
https://doi.org/10.1137/24M1634709

Funding: This work was supported by the DARPA Assured Autonomy and ANSR programs, the NASA ULI program in Safe Aviation Autonomy, and the ONR Basic Research Challenge in Multibody Control Systems. The fourth author was supported by the National Science Foundation under grant 2211548.

†Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720 USA (jingqili@berkeley.edu, sojoudi@berkeley.edu, tomlin@eecs.berkeley.edu).

‡Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, TX 78712-1221 USA (dfk@utexas.edu).

can compute an *open-loop Stackelberg equilibrium* (OLSE). At such an equilibrium, players' decisions depend on the initial state of a game and followers' decisions are influenced by the leaders'. When players also have access to state information and their prior players' actions, it becomes appropriate to compute a *feedback Stackelberg equilibrium* (FSE), where each player's decision is contingent upon the current state and the actions of preceding players. One advantage of FSE over OLSE is its subgame perfection, meaning that decision policies remain optimal for future stages, even if the state is perturbed at an intermediate stage. This feature is particularly beneficial in scenarios with feedback interactions among players, such as in lane-merging during highway driving [44] and human-robot interactions [14]. In these situations, the subgame perfection of FSE makes it a more suitable equilibrium concept than OLSE, as it allows players to adjust their decisions based on the current state information.

Though FSE is conceptually appealing, computing it poses significant challenges [18, 33, 49, 50]. Previous research has extensively explored the FSE problem in finite dynamic games, characterized by a finite number of states and actions [1, 5, 24, 42, 45, 48]. In contrast, infinite dynamic games—those with an infinite number of states and actions—have mostly been considered within the framework of linear quadratic (LQ) games, featuring linear dynamics and stagewise quadratic costs [5, 12, 15, 21, 46]. The computation of FSE for more general nonlinear games is more challenging than for LQ games. A naive application of existing dynamic programming solutions in finite dynamic games necessitates gridding the continuous state and action spaces, often leading to computational intractability [6]. Recent works [35, 51] have proposed using approximate dynamic programming to compute an approximate FSE for input-affine systems. Additionally, several iterative LQ approximation approaches have been proposed in [19, 22], but they lack convergence guarantees.

Moreover, existing approaches are ill-suited for handling coupled equality and inequality constraints on players' states and decisions, which frequently arise in safety-critical applications such as autonomous driving [43] and human-robot interaction [23]. For instance, existing iterative LQ game solvers [19, 22] cannot be directly integrated with the primal log barrier penalty method [39] to incorporate these constraints. The most relevant studies, such as [13, 32, 34], focus on computing OLSE in games under linear constraints. This paper aims to bridge this gap in the literature.

Our contributions are threefold: (1) We first reformulate the N-player FSE problem, characterized by N players making sequential decisions over time, into a sequence of nested optimization problems. This reformulation enables us to derive the Karush–Kuhn–Tucker (KKT) conditions and a second-order sufficient condition for the FSE. (2) Using these results, we propose a Newton-style primal-dual interior point (PDIP) algorithm for computing a local FSE for LQ games. Under certain regularity conditions, we show the convergence of our algorithm to a local FSE. (3) Finally, we propose an efficient PDIP method for approximately computing a local FSE for more general nonlinear games under (nonconvex) coupled equality and inequality constraints. The computed feedback policy locally approximates the ground truth nonlinear policy. Theoretically, we characterize the approximation error of our method and show the exponential convergence under certain conditions. Empirically, we validate our algorithm in a highway lane-merging scenario, demonstrating its ability to tolerate infeasible initializations and efficiently converge to a local FSE in constrained nonlinear games.

2. Related works. Closely related to the feedback Stackelberg equilibrium (FSE), the feedback Nash equilibrium (FNE) has been extensively studied, for

example, in [4, 5, 25, 40]. Our work builds upon [25], where the authors proposed KKT conditions for constrained FNE. However, the FNE KKT conditions in [25] fail to hold true for FSE due to the decision hierarchy in FSE. In our work, we introduce a set of new KKT conditions for FSE. Another key difference is that we adopt the PDIP method for solving LQ and nonlinear games, whereas [25] considers the active-set method. In general, the former has polynomial complexity, but the latter has exponential complexity [16]. Moreover, we are able to prove the exponential convergence of our algorithm under certain conditions. However, there is no such convergence proof in [25].

As highlighted in the literature, e.g., [5, 29, 35, 46, 51], the dominant approach to computing unconstrained FSE is using (approximate) dynamic programming. LQ games can be solved efficiently via exact dynamic programming; however, in more general nonlinear cases the value function could be hard to compute and, in general, has no analytical solution [35]. Compared with those works, our approach could be considered as computing an efficient local approximation of the value function along the state trajectory under a local FSE policy instead of approximating the value function everywhere as in [35].

Finally, to further motivate our work, we discuss whether the FSE could be approximated well by an FNE or an OLSE. As suggested by [24], the FSE could coincide with the FNE in repeated matrix games. However, we show a counterexample in Appendix A.1 that the FSE could be arbitrarily different from the FNE in LQ games. Moreover, there is a recent trend of approximating feedback policies via receding-horizon open-loop policies [27, 53], where an open-loop policy is re-solved at each time for future steps. However, we show in another counterexample in Appendix A.2 that the trajectory under the feedback Stackelberg policy and the one under the receding-horizon open-loop Stackelberg policy could be quite different, even if there is no state perturbation. Thus, it is essential to develop specific tools for computing the FSE.

3. Constrained feedback Stackelberg games. In this section, we introduce the formulation of constrained feedback Stackelberg games. We formulate the problem by extending the N-player feedback Stackelberg games [15] to its constrained setting. We denote by $\mathbb{N}$ and $\mathbb{R}$ the sets of natural numbers and real numbers, respectively. Given $j, k \in \mathbb{N}$, we denote by $\mathbf{I}_j^k = \{j, j+1, \ldots, k\}$ if $j \leq k$ and $\emptyset$ otherwise. Let $T \in \mathbb{N}$ be the time horizon over which the game is played. At each time t, we denote by x_t and $u_t^i \in \mathbb{R}^{m_i}$ the state of the entire game and the control input of player i, respectively. We define $u_t := [u_t^1, u_t^2, \ldots, u_t^N] \in \mathbb{R}^m$, with $m := \sum_{i=1}^N m_i$, to be the joint control input at time t. Moreover, at each time t, players make decisions in the order of their indices. We consider the time-varying dynamics

$$x_{t+1} = f_t(x_t, u_t), \tag{3.1}$$

where $f_t(x_t, u_t) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is assumed to be a twice differentiable function. Given a sequence of control inputs $\mathbf{u} := [u_0, u_1, \ldots, u_T] \in \mathbb{R}^{Tm}$, we denote by $\mathbf{x} := [x_0, x_1, \ldots, x_{T+1}] \in \mathbb{R}^{(T+1)n}$ a state trajectory under dynamics (3.1).

At each time $t \in \mathbf{I}_0^T$, we denote the stagewise cost of player $i \in \mathbf{I}_1^N$ by $\ell_t^i(x_t, u_t) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and associate with each player a terminal cost, $\ell_{T+1}^i(x_{T+1}) : \mathbb{R}^n \to \mathbb{R}$. Each player $i \in \mathbf{I}_1^N$ considers the following time-separable costs:

$$J^i(\mathbf{x}, \mathbf{u}) = \sum_{t=0}^{T} \ell_t^i(x_t, u_t) + \ell_{T+1}^i(x_{T+1}). \tag{3.2}$$

Moreover, let $n^i_{h,t}$ and $n^i_{g,t}$ be the number of equality and inequality constraints held by player $i \in \mathbf{I}_1^N$ at time t, respectively. We denote the equality and inequality constraint functions of player i by $h^i_t(x_t, u_t) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n^i_{h,t}}$ and $g^i_t(x_t, u_t) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n^i_{g,t}}$, respectively. We specify the stagewise equality and inequality constraints of player $i \in \mathbf{I}_1^N$ as

$$0 = h^i_t(x_t, u_t),\ 0 \le g^i_t(x_t, u_t). \tag{3.3}$$

At the terminal time $t = T + 1$, we represent the equality and inequality constraint functions of player $i \in \mathbf{I}_1^N$ by $h^i_{T+1}(x_{T+1}) : \mathbb{R}^n \to \mathbb{R}^{n^i_{h,T+1}}$ and $g^i_{T+1}(x_{T+1}) : \mathbb{R}^n \to \mathbb{R}^{n^i_{g,T+1}}$, respectively. We consider the following equality and inequality constraints of player $i \in \mathbf{I}_1^N$ at the terminal time,

$$0 = h^i_{T+1}(x_{T+1}),\ 0 \le g^i_{T+1}(x_{T+1}). \tag{3.4}$$

We remark that these definitions generate coupled dynamics and constraints among different players at each time $t \in \mathbf{I}_0^{T+1}$. We consider the following regularity assumption, following [11, 25].

Assumption 3.1. The feasible set $\mathcal{F} := \{x \in \mathbb{R}^{(T+1)n}, u \in \mathbb{R}^{Tm} : h^i_t(x_t, u_t) = 0, g^i_t(x_t, u_t) \ge 0, h^i_{T+1}(x_{T+1}) = 0, g^i_{T+1}(x_{T+1}) \ge 0, x_{t+1} = f_t(x_t, u_t)\ \forall i \in \mathbf{I}_1^N, t \in \mathbf{I}_0^T\}$ is compact. The costs, dynamics, and equality and inequality constraints are twice differentiable and bounded, but could be nonconvex in general.

3.1. Local feedback Stackelberg equilibria. In this subsection, we formalize the decision process of feedback Stackelberg games. Before doing that, we introduce a few notations to compactly represent different players' control at different times. We define $u^{i:i'}_{t:t'} := \{u^j_\tau, \tau \in \mathbf{I}_t^{t'}, j \in \mathbf{I}_i^{i'}\}$. In particular, we define $u_t^{1:i-1} := \emptyset$ when $i = 1$ and $u_t^{i+1:N} := \emptyset$ when $i = N$. We also denote by $u^{1:i}_{t+1:T} := \emptyset$ when $t = T$.

The policy of each player can be defined as follows. At the tth stage, since player 1 makes a decision first, its policy function $\pi^1_t(x_t) : \mathbb{R}^n \to \mathbb{R}^{m_1}$ depends only on the state x_t. For players $i \in \mathbf{I}_2^N$, the policies are modeled as $\pi^i_t(x_t, u_t^{1:i-1}) : \mathbb{R}^n \times \mathbb{R}^{\sum_{j=1}^{i-1} m_j} \to \mathbb{R}^{m_i}$. We will define the concept of *local feedback Stackelberg equilibria* in the remainder of this subsection.

At the terminal time $t = T + 1$, we define the state-value functions for a player $i \in \mathbf{I}_1^N$ as

$$V^i_{T+1}(x_{T+1}) := \begin{cases} \ell^i_{T+1}(x_{T+1}) & \text{if } \begin{cases} 0 = h^i_{T+1}(x_{T+1}), \\ 0 \le g^i_{T+1}(x_{T+1}), \end{cases} \\ \infty & \text{else.} \end{cases} \tag{3.5}$$

At time $t \le T$, we first construct the state-action-value function for player N:

$$Z^N_t(x_t, u_t^{1:N-1}, u_t^N) := \begin{cases} \ell^N_t(x_t, u_t) + V^N_{t+1}(x_{t+1}) & \text{if } \begin{cases} 0 = x_{t+1} - f_t(x_t, u_t), \\ 0 = h^N_t(x_t, u_t), \\ 0 \le g^N_t(x_t, u_t), \end{cases} \\ \infty & \text{else.} \end{cases} \tag{3.6}$$

Given $(x_t, u_t^{1:N-1})$, there could be multiple u_t^N minimizing $Z^N_t(x_t, u_t^{1:N-1}, u_t^N)$. We define player N's local FSE policy π^N_t by picking an arbitrary local minimizer u_t^{N*},

$$\pi^N_t(x_t, u_t^{1:N-1}) := u_t^{N*} \in \arg\min_{\tilde{u}_t^N} Z^N_t(x_t, u_t^{1:N-1}, \tilde{u}_t^N). \tag{3.7}$$

We then construct the state-action-value function of player $i \in \mathbf{I}_2^{N-1}$,

$$Z_t^i(x_t, u_t^{1:i-1}, u_t^i) := \begin{cases} \ell_t^i(x_t, u_t) + V_{t+1}^i(x_{t+1}) & \text{if } \begin{cases} 0 = x_{t+1} - f_t(x_t, u_t), \\ 0 = h_t^i(x_t, u_t), \\ 0 \le g_t^i(x_t, u_t), \\ u_t^j = \pi_t^j(x_t, u_t^{1:j-1}),\ j \in \mathbf{I}_{i+1}^N, \end{cases} \\ \infty & \text{else,} \end{cases} \tag{3.8}$$

and its local FSE policy π_t^i by picking an arbitrary local minimizer $u_t^{i^*}$,

$$\pi_t^i(x_t, u_t^{1:i-1}) := u_t^{i*} \in \arg\min_{\tilde{u}_t^i} Z_t^i(x_t, u_t^{1:i-1}, \tilde{u}_t^i). \tag{3.9}$$

We finally construct the state-action-value function of the first player,

$$Z_t^1(x_t, u_t^1) := \begin{cases} \ell_t^1(x_t, u_t) + V_{t+1}^1(x_{t+1}) & \text{if } \begin{cases} 0 = x_{t+1} - f_t(x_t, u_t), \\ 0 = h_t^1(x_t, u_t), \\ 0 \le g_t^1(x_t, u_t), \\ u_t^j = \pi_t^j(x_t, u_t^{1:j-1}),\ j \in \mathbf{I}_2^N, \end{cases} \\ \infty & \text{else,} \end{cases} \tag{3.10}$$

and its local FSE policy

$$\pi_t^1(x_t) := u_t^{1*} \in \arg\min_{\tilde{u}_t^1} Z_t^1(x_t, \tilde{u}_t^1). \tag{3.11}$$

We define the state-value function of player $i \in \{1, 2, \ldots, N\}$ at time $t \le T$ as

$$V_t^i(x_t) = Z_t^i(x_t, u_t^{1*}, \ldots, u_t^{i*}), \tag{3.12}$$

where $u_t^{j*} = \pi_t^j(x_t, u_t^{1:(j-1)*})\ \forall j \in \mathbf{I}_1^i$.

We formally define the local feedback Stackelberg equilibria as follows.

DEFINITION 3.2 (local feedback Stackelberg equilibria [5]). *Let $\{\pi_t^i\}_{t=0,i=1}^{T,N}$ be a set of policies defined in* (3.7), (3.9), *and* (3.11), *and define $(\mathbf{x}^*, \mathbf{u}^*)$ to be a state and control trajectory under the policies $\{\pi_t^i\}_{t=0,i=1}^{T,N}$, i.e.,*

$$x_{t+1}^* = f_t(x_t^*, u_t^*),\ u_t^{i*} = \pi_t^i(x_t^*, u_t^{1:(i-1)*})\ \ \forall t \in \mathbf{I}_0^T,\ i \in \mathbf{I}_1^N. \tag{3.13}$$

We say that $(\mathbf{x}^*, \mathbf{u}^*)$ is a local feedback Stackelberg equilibrium trajectory *if there exists an $\epsilon > 0$ such that, for all $t \in \mathbf{I}_0^T$,*

$$\begin{aligned} &Z_t^1(x_t^*, \tilde{u}_t^1) \ge Z_t^1(x_t^*, u_t^{1*}), \\ &\qquad\vdots \\ &Z_t^N(x_t^*, u_t^{1*}, \ldots, u_t^{(N-1)*}, \tilde{u}_t^N) \ge Z_t^N(x_t^*, u_t^{1*}, \ldots, u_t^{(N-1)*}, u_t^{N*}) \end{aligned} \tag{3.14}$$

for all $\tilde{u}_t^1 \in \{u : \|u - u_t^{1*}\|_2 \le \epsilon\}, \ldots$ and for all $\tilde{u}_t^N \in \{u : \|u - u_t^{N*}\|_2 \le \epsilon\}$.

The above definition encapsulates the traditional approach to computing feedback Stackelberg equilibria. This involves optimizing over state-action-value functions, which are obtained by integrating other players' policies into each player's problem and then recording the overall costs.

Remark 3.3 (existence of local feedback Stackelberg equilibria). In general, it is difficult to establish a sufficient condition for the existence of an FSE [7]. The main difficulty is that the decision problem of each player is nested within that of other

players. It must be solved hierarchically. For example, the existence of feedback Stackelberg policies [31] of a player $i \in \mathbf{I}_1^{N-1}$ is related to the topological properties of the set of policies of players $j \in \mathbf{I}_{i+1}^N$. Even if all the players' costs are convex, the feedback Stackelberg policy of player N at the terminal time could be lower semicontinuous. Subsequently, the cost of player $(N-1)$ could become upper semicontinuous when substituting in the Nth player's policy into the $(N-1)$th player's cost. Since there may not exist a solution when minimizing an upper semicontinuous function, there may not exist a feedback Stackelberg policy for player $(N-1)$. However, if we can show that the policy of each player is always continuous in the state and prior players' controls, and the continuous costs are defined on a compact domain, then there exist feedback Stackelberg equilibria [5].

We will now proceed to characterize the feedback Stackelberg equilibria in greater detail in the subsequent section.

4. Necessary and sufficient conditions for local feedback Stackelberg equilibria. We show in the following theorem that the dynamic programming problem, as described in Definition 3.2, can be reformulated as a sequence of nested constrained optimization problems. In this reformulation, the policies for other players are integrated as constraints within the problem of each player i, instead of being directly substituted into the costs for computing state-action-value functions, as is typical in traditional optimal control literature. This approach enables us to establish KKT conditions for feedback Stackelberg games in the latter part of this subsection.

THEOREM 4.1. *Under Assumption* 3.1*, for each* $t \in \mathbf{I}_0^T$ *and each* $i \in \mathbf{I}_1^N$*, a local feedback Stackelberg policy* π_t^i *can be equivalently represented as an optimization problem, given the knowledge of current state* $\bar{x}_t$ *and prior players' actions* $\bar{u}_t^{1:i-1}$*:*

(4.1a)
$$\pi_t^i(\bar{x}_t, \bar{u}_t^{1:i-1}) = \tilde{u}_t^i \in \arg \min_{\substack{u_t^i \\ u_t^{i:N} \\ u_{t+1:T}^{1:N} \\ x_{t+1:T+1}}} \ell_t^i(\bar{x}_t, \bar{u}_t^{1:i-1}, u_t^{i:N}) + \sum_{\tau=t+1}^{T} \ell_\tau^i(x_\tau, u_\tau) + \ell_{T+1}^i(x_{T+1})$$

(4.1b) $\quad s.t.\ 0 = u_t^j - \pi_t^j(\bar{x}_t, \bar{u}_t^{1:i-1}, u_t^{i:j-1}), \qquad j \in \mathbf{I}_{i+1}^N,$

(4.1c) $\quad 0 = x_{t+1} - f_t(\bar{x}_t, \bar{u}_t^{1:i-1}, u_t^{i:N}),$

(4.1d) $\quad 0 = h_t^i(\bar{x}_t, \bar{u}_t^{1:i-1}, u_t^{i:N}),\ 0 \le g_t^i(\bar{x}_t, \bar{u}_t^{1:i-1}, u_t^{i:N}),$

(4.1e) $\quad 0 = u_\tau^j - \pi_\tau^j(x_\tau, u_\tau^{1:j-1}), \qquad \tau \in \mathbf{I}_{t+1}^T, j \in \mathbf{I}_1^N \setminus \{i\},$

(4.1f) $\quad 0 = x_{\tau+1} - f_\tau(x_\tau, u_\tau), \qquad \tau \in \mathbf{I}_{t+1}^T,$

(4.1g) $\quad 0 = h_\tau^i(x_\tau, u_\tau),\ 0 \le g_\tau^i(x_\tau, u_\tau), \qquad \tau \in \mathbf{I}_{t+1}^T,$

(4.1h) $\quad 0 = h_{T+1}^i(x_{T+1}),\ 0 \le g_{T+1}^i(x_{T+1}),$

where we drop (4.1b) *when* $i = N$*, and we drop* (4.1e)*,* (4.1f)*, and* (4.1g) *when* $t = T$*. The notation* $\arg_u \min_{u,v}$ *represents that we minimize over* (u, v) *but only return* u *as an output.*

Proof. The proof can be found in Appendix A. ∎

In what follows, we will characterize the KKT conditions of the constrained optimization problems in (4.1). Before doing that, we first introduce Lagrange multipliers, which facilitate the formulation of Lagrangian functions for all players.

Let $t \in \mathbf{I}_0^T$ and $i \in \mathbf{I}_1^N$. We denote by $\lambda_t^i \in \mathbb{R}^n$ the Lagrange multiplier for the dynamics constraint $0 = x_{t+1} - f_t(x_t, u_t)$. Let $\mathbb{R}_{\geq 0}$ be the set of nonnegative real numbers. We define $\mu_t^i \in \mathbb{R}^{n_{h,t}^i}$ and $\gamma_t^i \in \mathbb{R}_{\geq 0}^{n_{g,t}^i}$ as the Lagrange multipliers for the constraints $0 = h_t^i(x_t, u_t)$ and $0 \leq g_t^i(x_t, u_t)$, respectively. When $t \leq T$, the constrained problem (4.1) of player $i < N$ considers the feedback interaction constraint $0 = u_t^j - \pi_t^j(x_t, u_t^{1:j-1})$, $j \in \mathbf{I}_{i+1}^N$. Thus, we associate those constraints with multipliers $\psi_t^i := [\psi_t^{i,i+1}, \psi_t^{i,i+2}, \ldots, \psi_t^{i,N}]$, where $\psi_t^{i,j} \in \mathbb{R}^{m_i}$. Moreover, when $t < T$, the constrained problem (4.1) of a player $i \leq N$ includes the feedback interaction constraints $0 = u_{\tau+1}^j - \pi_{\tau+1}^j(x_{\tau+1}, u_{\tau+1}^{1:j-1})$ for $\tau \geq t$ and $j \in \mathbf{I}_1^N \setminus \{i\}$. Thus, we associate those constraints with multipliers $\eta_t^i := [\eta_t^{i,1}, \ldots, \eta_t^{i,i-1}, \eta_t^{i,i+1}, \ldots, \eta_t^{i,N}]$, where $\eta_t^{i,j} \in \mathbb{R}^{m_j}$. Finally, we simplify the notation by defining $\lambda_t := [\lambda_t^1, \lambda_t^2, \ldots, \lambda_t^N]$, and define μ_t, γ_t, η_t, and ψ_t accordingly.

Subsequently, we define the Lagrangian functions of all the players. We first consider player $i \in \mathbf{I}_1^N$,

(4.2)
$$\begin{aligned} L_t^i(x_{t:t+1}, u_{t:t+1}, \lambda_t, \mu_t, \gamma_t, \eta_t, \psi_t) := {} & \ell_t^i(x_t, u_t) - \lambda_t^{i\top}(x_{t+1} - f_t(x_t, u_t)) \\ & - \mu_t^{i\top} h_t^i(x_t, u_t) - \gamma_t^{i\top} g_t^i(x_t, u_t) \\ & - \sum_{j \in \mathbf{I}_{i+1}^N} \psi_t^{i,j\top}(u_t^j - \pi_t^j(x_t, u_t^{1:j-1})) \\ & - \sum_{j \in \mathbf{I}_1^N \setminus \{i\}} \eta_t^{i,j\top}(u_{t+1}^j - \pi_{t+1}^j(x_{t+1}, u_{t+1}^{1:j-1})), \end{aligned}$$

where the right-hand side terms represent player i's cost, dynamics constraint, equality and inequality constraints, and constraints encoding the feedback interaction among players at the current and future time steps.

Furthermore, at the terminal time $t = T$, for player $i \in \mathbf{I}_1^N$, we consider

(4.3)
$$\begin{aligned} L_T^i(x_{T:T+1}, u_T, \lambda_T, \mu_{T:T+1}, \gamma_{T:T+1}, \psi_T) := {} & \ell_T^i(x_T, u_T) + \ell_{T+1}^i(x_T) \\ & - \lambda_T^{i\top}(x_{T+1} - f_T(x_T, u_T)) \\ & - \mu_T^{i\top} h_T^i(x_T, u_T) - \mu_{T+1}^{i\top} h_{T+1}^i(x_{T+1}) \\ & - \gamma_T^{i\top} g_T^i(x_T, u_T) - \gamma_{T+1}^{i\top} g_{T+1}^i(x_{T+1}) \\ & - \sum_{j \in \mathbf{I}_{i+1}^N} \psi_T^{i,j\top}(u_T^j - \pi_T^j(x_T, u_T^{1:j-1})) \end{aligned}$$

where the right-hand side terms represent player i's costs, dynamics constraint, equality and inequality constraints, and constraints encoding the feedback interaction among players at the terminal time T. Note that there is no more decision to be made at time $t = T + 1$, and therefore, there is no term representing the feedback interactions among players for future time steps in (4.3), which is different from (4.2).

For all time steps $t \in \mathbf{I}_0^T$ and players $i \in \mathbf{I}_1^N$, assuming the state x_t is given and each player $j < i$ has taken action u_t^j, we formulate the Lagrangian of the problem (4.1) of player i at the tth stage as

$$\begin{aligned} & \mathcal{L}_t^i(x_{t:T+1}, u_{t:T}, \lambda_{t:T}, \mu_{t:T+1}, \gamma_{t:T+1}, \eta_{t:T-1}, \psi_{t:T}) \\ & \qquad := \sum_{\tau=t}^{T-1} L_\tau^i(x_{\tau:\tau+1}, u_{\tau:\tau+1}, \lambda_\tau, \mu_\tau, \gamma_\tau, \eta_\tau, \psi_\tau) \\ & \qquad\quad + L_T^i(x_{T:T+1}, u_T, \lambda_T, \mu_{T:T+1}, \gamma_{T:T+1}, \psi_T), \end{aligned} \tag{4.4}$$

where for each $\tau \in \mathbf{I}_{t+1}^{T}$, the terms associated with ψ_τ in L_τ^i ensure constraints already addressed by the terms associated with $\eta_{\tau-1}$ in $L_{\tau-1}^i$ and can therefore be dropped when defining $\mathcal{L}_t^i$. We can concatenate the KKT conditions of each player at each stage, and summarize the overall KKT conditions for (4.1) in the following theorem.

THEOREM 4.2 (necessary nondition). *Under Assumption* 3.1*, let* $(\mathbf{x}^*, \mathbf{u}^*)$ *be a local FSE trajectory. Suppose that the linear independence constraint qualification (LICQ)* [37] *and strict complementarity condition* [9] *are satisfied at* $(\mathbf{x}^*, \mathbf{u}^*)$*. Furthermore, suppose* $\{\pi_t^i\}_{t=0,i=1}^{T,N}$ *is a set of local feedback Stackelberg policies and* π_t^i *is differentiable around* $(x_t^*, u_t^{1:(i-1)*})$ $\forall t \in \mathbf{I}_0^T$, $i \in \mathbf{I}_1^N$*. The KKT conditions of* (4.1) *can be formulated as for all* $i \in \mathbf{I}_1^N$, $t \in \mathbf{I}_0^T$,

(4.5)

$$
\begin{aligned}
0 &= \nabla_{u_t^i} \mathcal{L}_t^i(x_{t:T+1}^*, u_{t:T}^*, \lambda_{t:T}, \mu_{t:T+1}, \gamma_{t:T+1}, \eta_{t:T-1}, \psi_{t:T}), \\
0 &= \nabla_{x_\tau} \mathcal{L}_t^i(x_{t:T+1}^* u_{t:T}^*, \lambda_{t:T}, \mu_{t:T+1}, \gamma_{t:T+1}, \eta_{t:T-1}, \psi_{t:T}) && \forall \tau \in \mathbf{I}_{t+1}^{T+1}, \\
0 &= \nabla_{u_t^j} \mathcal{L}_t^i(x_{t:T+1}^*, u_{t:T}^*, \lambda_{t:T}, \mu_{t:T+1}, \gamma_{t:T+1}, \eta_{t:T-1}, \psi_{t:T}) && \forall j \in \mathbf{I}_{i+1}^N, \\
0 &= \nabla_{u_\tau^j} \mathcal{L}_t^i(x_{t:T+1}^*, u_{t:T}^*, \lambda_{t:T}, \mu_{t:T+1}, \gamma_{t:T+1}, \eta_{t:T-1}, \psi_{t:T}) && \forall j \in \mathbf{I}_1^N \setminus \{i\}, \forall \tau \in \mathbf{I}_{t+1}^T, \\
0 &= x_{\tau+1}^* - f_\tau(x_\tau^*, u_\tau^*) && \forall \tau \in \mathbf{I}_t^T, \\
0 &= h_\tau^i(x_\tau^*, u_\tau^*) && \forall \tau \in \mathbf{I}_t^T, \\
0 &\le \gamma_\tau^i \perp g_\tau^i(x_\tau^*, u_\tau^*) \ge 0 && \forall \tau \in \mathbf{I}_t^T, \\
0 &= h_{T+1}^i(x_{T+1}^*), \\
0 &\le \gamma_{T+1}^i \perp g_{T+1}^i(x_{T+1}^*) \ge 0,
\end{aligned}
$$

where $\perp$ *represents the complementary slackness condition* [9]*. Then, there exist Lagrange multipliers* $\boldsymbol{\lambda} := [\lambda_t]_{t=0}^T$, $\boldsymbol{\mu} := [\mu_t]_{t=0}^T$, $\boldsymbol{\gamma} := [\gamma_t]_{t=0}^T$, $\boldsymbol{\eta} := [\eta_t]_{t=0}^{T-1}$, *and* $\boldsymbol{\psi} := [\psi_t]_{t=0}^T$*, such that* (4.5) *holds true.*

Proof. The proof can be found in Appendix A. ∎

Constructing the KKT conditions in (4.5) requires the computation of policy gradients, $\{\nabla \pi_t^i\}_{t=0,i=1}^{T,N}$, which appear in the first four rows of (4.5). However, knowing the policy itself is not required, as any solution satisfying the KKT conditions obeys the corresponding feedback Stackelberg policy, as shown in the proof of Theorem 4.2. A key distinction between (4.5) and the FNE KKT conditions in [25] lies in the accommodation of a decision hierarchy among the N players at each stage. This is reflected in the terms $-\sum_{j \in \mathbf{I}_{i+1}^N} \psi_t^{i,j\top}(u_t^j - \pi_t^j(x_t, u_t^{1:j-1}))$ in the Lagrangian $\mathcal{L}_t^i$. Additionally, this decision hierarchy differentiates the construction of the FSE KKT conditions from those of FNE. We will outline a detailed procedure for constructing the FSE KKT conditions in sections 5 and 6, with an example provided in Appendix B.

Furthermore, we propose a sufficient condition for FSE trajectories in the following theorem.

THEOREM 4.3 (sufficient condition). *Let* $(\mathbf{x}^*, \mathbf{u}^*)$ *be a trajectory and* $\{\pi_t^i\}_{t=0,i=1}^{T,N}$ *be the associated policies. Suppose there exist Lagrange multipliers* $\{\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\psi}\}$ *satisfying* (4.5) *and there exists an* $\epsilon > 0$ *such that, for all* $i \in \mathbf{I}_1^N$, $t \in \mathbf{I}_0^T$*, and nonzero*

$\{\Delta x_{T+1}\} \bigcup \{\Delta x_\tau, \Delta u_\tau\}_{\tau=t}^T$ *satisfying*

$$
\begin{aligned}
0 &= \Delta u_t^j - \nabla \pi_t^j(x_t^*, u_t^{1:(j-1)*}) \begin{bmatrix} \Delta x_t \\ \Delta u_t^{1:j-1} \end{bmatrix} \quad \forall j \in \mathbf{I}_i^N, \\
0 &= \Delta u_\tau^j - \nabla \pi_\tau^j(x_\tau^*, u_\tau^{1:(j-1)*}) \begin{bmatrix} \Delta x_\tau \\ \Delta u_\tau^{1:j-1} \end{bmatrix} \quad \forall j \in \mathbf{I}_1^N, \forall \tau \in \mathbf{I}_{t+1}^T, \\
0 &= \Delta x_{\tau+1} - \nabla f_\tau(x_\tau^*, u_\tau^*) \begin{bmatrix} \Delta x_\tau, \\ \Delta u_\tau \end{bmatrix} \quad \forall \tau \in \mathbf{I}_t^T, \\
0 &= \nabla h_\tau^j(x_\tau^*, u_\tau^*) \begin{bmatrix} \Delta x_\tau \\ \Delta u_\tau \end{bmatrix}, \quad 0 = \nabla h_{T+1}^j(x_{T+1}^*) \Delta x_{T+1} \quad \forall \tau \in \mathbf{I}_0^T, \forall j \in \mathbf{I}_1^N,
\end{aligned} \tag{4.6}
$$

we have $\sum_{\tau=t}^T [\begin{smallmatrix} \Delta x_\tau \\ \Delta u_\tau^i \end{smallmatrix}]^\top \nabla^2_{[x_\tau^*, u_\tau^{i*}]} L_\tau^i [\begin{smallmatrix} \Delta x_\tau \\ \Delta u_\tau^i \end{smallmatrix}] + \Delta x_{T+1}^\top \nabla^2_{x_{T+1}^*} L_{T+1}^i \Delta x_{T+1} > 0$.

Then, $(\mathbf{x}^*, \mathbf{u}^*)$ *constitutes a local FSE trajectory.*

Proof. The proof can be found in Appendix A. ∎

Remark 4.4. The gap between the necessity condition in Theorem 4.2 and the sufficiency condition in Theorem 4.3 is due to the fact that a solution to (4.5) may not necessarily be an FSE, and that there exist feedback Stackelberg equilibria where the cost functions possess zero second-order gradients.

Theorems 4.2 and 4.3 establish conditions to certify whether a trajectory $(\mathbf{x}, \mathbf{u})$ constitutes an FSE with a set of feedback Stackelberg policies $\{\pi_t^i\}_{t=0,i=1}^{T,N}$. However, computing feedback Stackelberg equilibria can be challenging. In the following sections, we will discuss how to approximately compute local feedback Stackelberg equilibria. We will first compute feedback Stackelberg equilibria for LQ games and then extend the result to nonlinear games.

5. Constrained LQ games.

We consider the linear dynamics

$$
x_{t+1} = f_t(x_t, u_t) = A_t x_t + B_t^1 u_t^1 + \cdots + B_t^N u_t^N + c_t, \ t \in \mathbf{I}_0^T, \tag{5.1}
$$

where $A_t \in \mathbb{R}^{n \times n}$, $B_t^i \in \mathbb{R}^{n \times m_i}$, and $c_t \in \mathbb{R}^n$. We denote $B_t := [B_t^1, B_t^2, \ldots, B_t^N]$. The cost of the ith player is defined as

$$
\begin{aligned}
\ell_t^i(x_t, u_t) &= \frac{1}{2} \begin{bmatrix} x_t \\ u_t \end{bmatrix}^\top \begin{bmatrix} Q_t^i & S_t^{i\top} \\ S_t^i & R_t^i \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix} + q_t^{i\top} x_t + r_t^{i\top} u_t, \ t \in \mathbf{I}_0^T, \\
\ell_{T+1}^i(x_{T+1}) &= \frac{1}{2} x_{T+1}^\top Q_{T+1}^i x_{T+1} + q_{T+1}^{i\top} x_{T+1},
\end{aligned} \tag{5.2}
$$

where symmetric matrices $Q_t^i \in \mathbb{R}^{n \times n}$ and $R_t^i \in \mathbb{R}^{m \times m}$ are positive semidefinite and positive definite, respectively. The off-diagonal matrix is denoted as $S_t^i \in \mathbb{R}^{m \times n}$. In particular, we partition the structure of R_t^i, S_t^i and r_t^i as follows:

$$
R_t^i = \begin{bmatrix} R_t^{i,1,1} & R_t^{i,1,2} & \cdots & R_t^{i,1,N} \\ R_t^{i,2,1} & R_t^{i,2,2} & \cdots & R_t^{i,2,N} \\ \vdots & \vdots & \ddots & \vdots \\ R_t^{i,N,1} & R_t^{i,N,2} & \cdots & R_t^{i,N,N} \end{bmatrix}, \quad S_t^i = \begin{bmatrix} S_t^{i,1} \\ S_t^{i,2} \\ \vdots \\ S_t^{i,N} \end{bmatrix}, \quad r_t^i = \begin{bmatrix} r_t^{i,1} \\ r_t^{i,2} \\ \vdots \\ r_t^{i,N} \end{bmatrix}, \tag{5.3}
$$

where $R_t^{i,j,k}$, $S_t^{i,j}$, and r_t^i represent the cost terms $u_t^{j\top} R_t^{i,j,k} u_t^k$, $u_t^{j\top} S_t^{i,j} x_t$, and $r_t^{i,j^\top} u_t^j$ in $\ell_t^i(x_t, u_t)$. The linear equality and inequality constraints are specified as

$$
\begin{aligned}
0 &= h_t^i(x_t, u_t) = H_{x_t}^i x_t + \sum_{j \in \mathbf{I}_1^N} H_{u_t^j}^i u_t^j + \bar{h}_t^i, \quad t \in \mathbf{I}_0^T, \\
0 &\le g_t^i(x_t, u_t) = G_{x_t}^i x_t + \sum_{j \in \mathbf{I}_1^N} G_{u_t^j}^i u_t^j + \bar{g}_t^i, \quad t \in \mathbf{I}_0^T, \\
0 &= h_{T+1}^i(x_{T+1}) = H_{x_{T+1}}^i x_{T+1} + \bar{h}_{T+1}^i, \\
0 &\le g_{T+1}^i(x_{T+1}) = G_{x_{T+1}}^i x_{T+1} + \bar{g}_{T+1}^i.
\end{aligned} \tag{5.4}
$$

5.1. Computing feedback Stackelberg equilibria and constructing the KKT conditions for LQ games. In this subsection, we introduce a process for deriving FSE and the KKT conditions for LQ games. When we have linear inequality constraints, the optimal policies of LQ games are generally piecewise linear functions of the state [8, 25]. However, this makes them nondifferentiable at the facets. In our work, we propose to use the primal-dual interior point (PDIP) method [37] to solve constrained LQ games. The benefits of using PDIP are its polynomial complexity and tolerance of infeasible initializations. Critically, under certain conditions, PDIP yields a local differentiable policy approximation to the ground truth piecewise linear policy, as shown in the rest of this section and an example in Appendix A.3.

To this end, we introduce a set of nonnegative slack variables $\{s_t^i\}_{t=0,i=1}^{T+1,N}$ such that we can rewrite the inequality constraints as equality constraints for $t \in \mathbf{I}_0^{T+1}$ and $i \in \mathbf{I}_1^N$,

$$
g_t^i(x_t, u_t) - s_t^i = 0, \quad g_{T+1}^i(x_{T+1}) - s_{T+1}^i = 0. \tag{5.5}
$$

In this paper, we consider PDIP as a homotopy method as in [37]. Instead of solving the mixed complementarity problem (4.5) directly, we seek solutions to the homotopy approximation of the complementary slackness condition

$$
\gamma_t^i \odot s_t^i = \rho \mathbf{1}, \quad s_t^i \ge 0, \quad \gamma_t^i \ge 0, \tag{5.6}
$$

where $\odot$ denotes the elementwise product and $\rho > 0$ is a hyperparameter to be reduced to 0 gradually such that we recover the ground truth solution when $\rho \to 0$. In the following section, we will construct the KKT conditions where we replace the mixed complementarity condition with its approximation (5.6). For each $\rho > 0$, we denote its corresponding local feedback policy as $\{\pi_{t,\rho}^i\}_{t=0,i=1}^{T,N}$, if it exists.

As shown in Theorem 4.2, the construction of the KKT conditions for player i at stage t requires the policy gradients of subsequent players at the current stage and future stages. In what follows, we construct those KKT conditions in reverse player order and backward in time.

5.1.1. Player N at the Tth stage. Before constructing the KKT conditions, we first introduce the variables of player N at the terminal time T, $\mathbf{z}_T^N := [u_T^N, \lambda_T^N, \mu_{T:T+1}^N, \gamma_{T:T+1}^N, s_{T:T+1}^N, x_{T+1}]$. As shown in Theorem 4.2, the KKT condi-

tions of player N at time T can be written as

$$0 = K^N_{T,\rho}(\mathbf{z}^N_T) := \begin{bmatrix} \nabla_{u^N_T} L^N_T \\ \nabla_{x_{T+1}} L^N_T \\ x_{T+1} - f_T(x_T, u_T) \\ h^N_T(x_T, u_T) \\ h^N_{T+1}(x_{T+1}) \\ g^N_T(x_T, u_T) - s^N_T \\ g^N_{T+1}(x_{T+1}) - s^N_{T+1} \\ \gamma^N_{T:T+1} \odot s^N_{T:T+1} - \rho \mathbf{1} \end{bmatrix}, \tag{5.7}$$

where the rows of $K^N_{T,\rho}(\mathbf{z}^N_T)$ represent the stationarity conditions with respect to u^N_T and x_{T+1}, the dynamics constraint, equality constraints, inequality constraints, and relaxed complementarity conditions. To obtain a local policy and its policy gradient around a $\mathbf{z}^N_T$ satisfying (5.7), we build a first-order approximation to (5.7),

$$\nabla K^N_{T,\rho} \cdot \Delta \mathbf{z}^N_T + \nabla_{[x_T, u^{1:N-1}_T]} K^N_{T,\rho} \cdot \begin{bmatrix} \Delta x_T \\ \Delta u^{1:N-1}_T \end{bmatrix} + K^N_{T,\rho}(\mathbf{z}^N_T) = 0. \tag{5.8}$$

If there is no solution $\Delta \mathbf{z}^N_T$ to (5.8), then we claim there is no feedback Stackelberg policy. Suppose (5.8) has a solution $\Delta \mathbf{z}^N_T$; then we can define $\Delta \mathbf{z}^N_T$ as

$$\Delta \mathbf{z}^N_T = \underbrace{-\left(\nabla K^N_{T,\rho}\right)^+ \cdot \left(\nabla_{[x_T, u^{1:N-1}_T]} K^N_{T,\rho} \cdot \begin{bmatrix} \Delta x_T \\ \Delta u^{1:N-1}_T \end{bmatrix} + K^N_{T,\rho}(\mathbf{z}^N_T) \right)}_{F^N_T(\Delta x_T, \Delta u^{1:N-1}_T)}, \tag{5.9}$$

where $(\cdot)^+$ represents the pseudoinverse and we denote $\Delta \mathbf{z}^N_T$ as a function F^N_T of $(\Delta x_T, \Delta u^{1:N-1}_T)$. Since Δu^N_T represents the first m_N entries of $\Delta \mathbf{z}^N_T$, we consider Δu^N_T as a function of $(\Delta x_T, \Delta u^{1:N-1}_T)$,

$$\Delta u^N_T = -\left[\left(\nabla K^N_{T,\rho}\right)^+\right]_{u^N_T} \cdot \left(\nabla_{[x_T, u^{1:N-1}_T]} K^N_{T,\rho} \cdot \begin{bmatrix} \Delta x_T \\ \Delta u^{1:N-1}_T \end{bmatrix} + K^N_{T,\rho}(\mathbf{z}^N_T) \right), \tag{5.10}$$

where $\left[\left(\nabla K^N_{T,\rho}\right)^+\right]_{u^N_T}$ represents the rows of the matrix $\left(\nabla K^N_{T,\rho}\right)^+$ corresponding to the variable u^N_T, i.e., the first m_N rows of the matrix $\left(\nabla K^N_{T,\rho}\right)^+$.

Furthermore, for some $x \in \mathbb{R}^n$ and $u^{1:N-1} \in \mathbb{R}^{\sum_{i=1}^{N-1} m_i}$, let $\Delta x_T = x - x_T$, $\Delta u^{1:N-1}_T = u^{1:N-1} - u^{1:N-1}_T$, and $\Delta u^N_T = u^N - u^N_T$. Substituting them into (5.10), we obtain a local policy $\tilde{\pi}^N_{T,\rho}$ for player N at time T,

$$\begin{aligned} u^N &= \tilde{\pi}^N_{T,\rho}(x, u^{1:N-1}) \\ &:= u^N_T - \left[\left(\nabla K^N_{T,\rho}\right)^+\right]_{u^N_T} \cdot \left(\nabla_{[x_T, u^{1:N-1}_T]} K^N_{T,\rho} \cdot \begin{bmatrix} x - x_T \\ u^{1:N-1} - u^{1:N-1}_T \end{bmatrix} + K^N_{T,\rho}(\mathbf{z}^N_T) \right). \end{aligned} \tag{5.11}$$

Suppose that $\nabla K^N_{T,\rho}(\mathbf{z}^N_T)$ has a constant row rank in an open set containing $\mathbf{z}^N_T$; then, by the constant rank theorem [20], the policy $\tilde{\pi}^N_{T,\rho}$ of player N at time T is locally differentiable with respect to $(x, u^{1:N-1})$, and its gradient over $(x, u^{1:N-1})$ is

$$\nabla \tilde{\pi}^N_{T,\rho} = -\left[\left(\nabla K^N_{T,\rho}\right)^+\right]_{u^N_T} \cdot \nabla_{[x_T, u^{1:N-1}_T]} K^N_{T,\rho}. \tag{5.12}$$

In the following subsection, we construct the KKT conditions of a player $i < N$ at stage T.

5.1.2. Players $i < N$ at the Tth stage. For player $i < N$, assuming that $\mathbf{z}_T^{i+1}$ has been defined and $\nabla\tilde{\pi}_{T,\rho}^{i+1}$ has been computed, we first introduce variables

$$\mathbf{y}_T^i := [u_T^i, \psi_T^i, \lambda_T^i, \mu_{T:T+1}^i, \gamma_{T:T+1}^i, s_{T:T+1}^i] \quad \text{and} \quad \mathbf{z}_T^i := [\mathbf{y}_T^i, \mathbf{z}_T^{i+1}]. \tag{5.13}$$

The KKT conditions of player i at time T are

$$0 = K_{T,\rho}^i(\mathbf{z}_T^i) := \begin{bmatrix} \hat{K}_{T,\rho}^i(\mathbf{y}_T^i) \\ K_{T,\rho}^{i+1}(\mathbf{z}_T^{i+1}) \end{bmatrix}, \quad \hat{K}_{T,\rho}^i(\mathbf{y}_T^i) := \begin{bmatrix} \nabla_{u_T^i} L_T^i \\ \nabla_{x_{T+1}} L_T^i \\ \nabla_{u_T^j} L_T^i \quad \forall j \in \mathbf{I}_{i+1}^N \\ h_T^i(x_T, u_T) \\ h_{T+1}^i(x_{T+1}) \\ g_T^i(x_T, u_T) - s_T^i \\ g_{T+1}^i(x_{T+1}) - s_{T+1}^i \\ \gamma_{T:T+1}^i \odot s_{T:T+1}^i - \rho\mathbf{1} \end{bmatrix}, \tag{5.14}$$

where the definition of L_T^i involves the policy $\pi_{T,\rho}^{i+1}$, as shown in (4.3). Building a first-order approximation to $0 = K_{T,\rho}^i(\mathbf{z}_T^i)$, we have

$$\nabla K_{T,\rho}^i \cdot \Delta\mathbf{z}_T^i + \nabla_{[x_T, u_T^{1:i-1}]} K_{T,\rho}^i \cdot \begin{bmatrix} \Delta x_T \\ \Delta u_T^{1:i-1} \end{bmatrix} + K_{T,\rho}^i(\mathbf{z}_T^i) = 0. \tag{5.15}$$

However, a drawback of PDIP is that the policy $\pi_{T,\rho}^{i+1}$ is nonlinear in state x_T and prior players' controls $u_T^{1:i}$, as shown in a simplified problem in Appendix A.3. The computation of $\nabla K_{T,\rho}^i$ involves the evaluation of $\nabla(\nabla_{u_T^i} L_T^i)$, which requires the computation of $\nabla(\psi_T^{i,i+1}\nabla\pi_{T,\rho}^{i+1}) = \nabla\psi_T^{i,i+1} \cdot \nabla\pi_{T,\rho}^{i+1} + \psi_T^{i,i+1} \cdot \nabla^2\pi_{T,\rho}^{i+1}$. Furthermore, to evaluate $\nabla^2\pi_{T,\rho}^{i+1}$, we need the computation of $\nabla^3\pi_{T,\rho}^{i+2}$. In other words, the construction of $\nabla K_{T,\rho}^i$ needs the evaluation of $\nabla^2\pi_{T,\rho}^{i+1}$, $\nabla^3\pi_{T,\rho}^{i+2}, \ldots, \nabla^{N-i+1}\pi_{T,\rho}^N$. The evaluation of high-order policy gradients is challenging in practice [25] because there is no closed-form solution to the KKT equation $0 = K_{T,\rho}^{i+1}(\mathbf{z}_T^{i+1})$.

We prove in Appendix A.3 that the high-order policy gradients could decay to zero as $\rho \to 0$, when the ground truth policy is piecewise linear and differentiable around x_T. Motivated by this observation, we propose to approximate the nonlinear policy $\pi_{T,\rho}^{i+1}$ by its first-order approximation $\tilde{\pi}_{T,\rho}^{i+1}$ in (5.11). With this approximation, we have $\nabla(\psi_T^{i,i+1}\nabla\tilde{\pi}_{T,\rho}^{i+1}) = \nabla\psi_T^{i,i+1} \cdot \nabla\tilde{\pi}_{T,\rho}^{i+1}$. We refer to this policy $\tilde{\pi}_{T,\rho}^{i+1}$ as a *quasi-policy.*

In the remainder of this section, we will always approximate the ground truth nonlinear policy by quasi-policy when we define the KKT conditions.

Solving (5.15), we can obtain $\Delta\mathbf{z}_T^i$ and $\nabla\tilde{\pi}_{T,\rho}^i$ as in (5.9) and (5.12), respectively. However, by construction, the dimension of $\Delta\mathbf{z}_T^i$ is higher than $\Delta\mathbf{z}_T^{i+1}$. Therefore, it is more expensive to compute $(\nabla K_{T,\rho}^i)^+$ than $(\nabla K_{T,\rho}^{i+1})^+$, and it is worthwhile to reduce the complexity of computing $\Delta\mathbf{z}_T^i$ by leveraging the computation that we have done for $\Delta\mathbf{z}_T^{i+1}$ and $\nabla\tilde{\pi}_{T,\rho}^{i+1}$. To this end, by exploiting the structure $\mathbf{z}_T^i = [\mathbf{y}_T^i, \mathbf{z}_T^{i+1}]$ in (5.13), we can rewrite (5.15) as

$$\begin{cases} \nabla\hat{K}_{T,\rho}^i \cdot \Delta\mathbf{y}_T^i + \nabla_{[x_T, u_T^{1:i-1}]} \hat{K}_{T,\rho}^i \cdot \begin{bmatrix} \Delta x_T \\ \Delta u_T^{1:i-1} \end{bmatrix} + \hat{K}_{T,\rho}^i(\mathbf{y}_T^i) = 0, \\ \nabla K_{T,\rho}^{i+1} \cdot \Delta\mathbf{z}_T^{i+1} + \nabla_{[x_T, u_T^{1:i}]} K_{T,\rho}^{i+1} \cdot \begin{bmatrix} \Delta x_T \\ \Delta u_T^{1:i} \end{bmatrix} + K_{T,\rho}^{i+1}(\mathbf{z}_T^{i+1}) = 0. \end{cases} \tag{5.16}$$

Observe that we have solved the second equation of (5.16) in section 5.1.1. What remains to be solved is the first equation in (5.16). We solve it as follows:

$$
\begin{aligned}
\Delta \mathbf{y}_T^i &= \underbrace{-\left(\nabla \hat{K}_{T,\rho}^i\right)^+ \cdot \left(\nabla_{[x_T, u_T^{1:i-1}]} \hat{K}_{T,\rho}^i \cdot \begin{bmatrix} \Delta x_T \\ \Delta u_T^{1:i-1} \end{bmatrix} + \hat{K}_{T,\rho}^i(\mathbf{y}_T^i)\right)}_{\hat{F}_T^i(\Delta x_T, \Delta u_T^{1:i-1})}, \\
\Delta u_T^i &= -\left[\left(\nabla \hat{K}_{T,\rho}^i\right)^+\right]_{u_T^i} \cdot \left(\nabla_{[x_T, u_T^{1:i-1}]} \hat{K}_{T,\rho}^i \cdot \begin{bmatrix} \Delta x_T \\ \Delta u_T^{1:i-1} \end{bmatrix} + \hat{K}_{T,\rho}^i(\mathbf{y}_T^i)\right), \\
\nabla \tilde{\pi}_{T,\rho}^i &= -\left[\left(\nabla \hat{K}_{T,\rho}^i\right)^+\right]_{u_T^i} \cdot \nabla_{[x_T, u_T^{1:i-1}]} \hat{K}_{T,\rho}^i.
\end{aligned} \tag{5.17}
$$

Combining (5.17) and (5.9), we have

$$
\Delta \mathbf{z}_T^i = \begin{bmatrix} \Delta \mathbf{y}_T^i \\ \Delta \mathbf{z}_T^{i+1} \end{bmatrix} = \begin{bmatrix} \hat{F}_T^i(\Delta x_T, \Delta u_T^{1:i-1}) \\ F_T^{i+1}(\Delta x_T, \Delta u_T^{1:i}) \end{bmatrix}. \tag{5.18}
$$

Since Δu_T^i is also a function of $(\Delta x_T, \Delta u_T^{1:i-1})$, as shown in (5.17), we can represent (5.18) compactly as $\Delta \mathbf{z}_T^i = F_T^i(\Delta x_T, \Delta u_T^{1:i-1})$.

As such, given that the KKT conditions of player $(i+1)$ at time T have been constructed, we have finished the construction of the KKT conditions for player i at time T, and we have introduced a computationally efficient way to compute $\nabla \tilde{\pi}_{T,\rho}^i$. We can derive the KKT conditions and quasi-policy gradient of player $i < N$ at time T, sequentially, from $i = N-1$ to $i = 1$.

5.1.3. Player N at a stage $t < T$. At a stage $t < T$, assuming that we have constructed the KKT conditions $0 = K_{t+1,\rho}^1(\mathbf{z}_{t+1}^1)$, we are ready to derive the KKT conditions for player N at time t. We first introduce the variable $\mathbf{z}_t^N := [\mathbf{y}_t^N, \mathbf{z}_{t+1}^1]$, with $\mathbf{y}_t^N := [u_t^N, \eta_t^N, \lambda_t^N, \mu_t^N, \gamma_t^N, s_t^N, x_{t+1}]$. We construct the KKT conditions of player N at time t as follows:

$$
0 = K_{t,\rho}^N(\mathbf{z}_t^N) := \begin{bmatrix} \nabla_{u_t^N} L_t^N \\ \nabla_{x_{t+1}} L_t^N \\ \nabla_{u_{t+1}^j} L_t^N \quad \forall j \in \mathbf{I}_1^{N-1} \\ x_{t+1} - f_t(x_t, u_t) \\ h_t^N(x_t, u_t) \\ g_t^N(x_t, u_t) - s_t^N \\ \gamma_t^N \odot s_t^N - \rho \mathbf{1} \\ K_{t+1,\rho}^1(\mathbf{z}_{t+1}^1) \end{bmatrix}. \tag{5.19}
$$

Building a first-order approximation to the above equation, we can obtain quasi-policy gradient $\nabla \tilde{\pi}_{t,\rho}^N$ as in (5.17) when it exists.

5.1.4. Players $i < N$ at a stage $t < T$. Suppose that we have constructed the KKT conditions for the $(i+1)$th player at the tth stage; we are then ready to construct the KKT conditions for player i at the tth stage. We introduce the variable $\mathbf{z}_t^i := [\mathbf{y}_t^i, \mathbf{z}_t^{i+1}]$ with $\mathbf{y}_t^i := [u_t^i, \psi_t^i, \eta_t^i, \lambda_t^i, \mu_t^i, \gamma_t^i, s_t^i]$. The KKT conditions of player i at time t are

Algorithm 5.1. Local Feedback Stackelberg Equilibrium via PDIP.

Require: $\{f_t\}_{t=0}^{T}$, $\{\ell_t^i, h_t^i, g_t^i\}_{t=0,i=1}^{T+1,N}$, initial homotopy parameter ρ, contraction rate $\sigma \in (0,1)$, parameters $\beta \in (0,1)$ and $\kappa \in (0,1)$, tolerance ϵ, initial solution $\mathbf{z}_\rho^{(0)} := [\mathbf{x}_\rho^{(0)}, \mathbf{u}_\rho^{(0)}, \boldsymbol{\lambda}_\rho^{(0)}, \boldsymbol{\mu}_\rho^{(0)}, \boldsymbol{\gamma}_\rho^{(0)}, \boldsymbol{\eta}_\rho^{(0)}, \boldsymbol{\psi}_\rho^{(0)}, \mathbf{s}_\rho^{(0)}]$ with $\mathbf{s}_\rho^{(0)} > 0$ and $\boldsymbol{\gamma}_\rho^{(0)} > 0$

Ensure: policies $\{\tilde{\pi}_{t,\rho}^i\}_{t=0,i=1}^{T,N}$, converged solution $\mathbf{z}_\rho$

for $k^{\text{out}} = 1, 2, \ldots, k_{\max}^{\text{out}}$ **do**
  while the merit function $\|K_\rho(\mathbf{z}_\rho^{(k)})\|_2 > \epsilon$ **do**
    construct the first-order approximation of the KKT conditions $0 = \nabla K_\rho \cdot \Delta \mathbf{z}_\rho + K_\rho(\mathbf{z}_\rho^{(k)})$
    $\Delta \mathbf{z}_\rho \leftarrow -\big(\nabla K_\rho\big)^+ \cdot K_\rho(\mathbf{z}_\rho^{(k)})$
    initialize the step size for line search, $\alpha \leftarrow 1$
    while $\|K_\rho(\mathbf{z}_\rho^{(k)} + \alpha \Delta \mathbf{z}_\rho)\|_2 > \kappa \|K_\rho(\mathbf{z}_\rho^{(k)})\|_2$ or $\hat{\mathbf{z}}_\rho := (\mathbf{z}_\rho^{(k)} + \alpha \Delta \mathbf{z}_\rho)$ has a nonpositive element in its subvector $[\hat{\mathbf{s}}_\rho, \hat{\boldsymbol{\gamma}}_\rho]$ **do**
      $\alpha \leftarrow \beta \cdot \alpha$
    end while
    if $\alpha == 0$ **then**
      claim **failure** to find a feedback Stackelberg equilibrium
    end if
    $\mathbf{z}_\rho^{(k+1)} \leftarrow \mathbf{z}_\rho^{(k)} + \alpha \Delta \mathbf{z}_\rho$
  end while
  $\rho \leftarrow \sigma \cdot \rho$
end for
construct $\{\tilde{\pi}_{t,\rho}^i\}_{t=0,i=1}^{T,N}$ as in (5.11) and record $\mathbf{z}_\rho \leftarrow \mathbf{z}_\rho^{(k)}$.
return $\{\tilde{\pi}_{t,\rho}^i\}_{t=0,i=1}^{T,N}$, $\mathbf{z}_\rho$

$$0 = K_{t,\rho}^i(\mathbf{z}_t^i) := \begin{bmatrix} \nabla_{u_t^i} L_t^i \\ \nabla_{x_{t+1}} L_t^i \\ \nabla_{u_t^j} L_t^i \quad \forall j \in \mathbf{I}_{i+1}^N \\ \nabla_{u_{t+1}^j} L_t^i \quad \forall j \in \mathbf{I}_1^N \setminus \{i\} \\ h_t^i(x_t, u_t) \\ g_t^i(x_t, u_t) - s_t^i \\ \gamma_t^i \odot s_t^i - \rho \mathbf{1} \\ K_{t,\rho}^{i+1}(\mathbf{z}_t^{i+1}) \end{bmatrix}. \tag{5.20}$$

Building a first approximation to the above equation, we can obtain the quasi-policy gradient $\nabla \tilde{\pi}_{t,\rho}^i$ as in (5.17), when it exists.

We observe that, by construction, the KKT conditions in (4.5) are equivalent to $0 = K_{0,\rho}^1(\mathbf{z}_0^1)$. To simplify notation, we define

$$\mathbf{z} := \mathbf{z}_0^1, \quad K_\rho(\mathbf{z}) := K_{0,\rho}^1(\mathbf{z}). \tag{5.21}$$

The KKT conditions (4.5) can be represented compactly as $0 = K_\rho(\mathbf{z})$. To more effectively illustrate the construction process of KKT conditions described above, we have included detailed examples of the KKT conditions for two-player LQ games in Appendix B as a reference.

5.2. PDIP algorithm and convergence analysis in constrained LQ games. In this subsection, we propose the application of Newton's method to compute $\mathbf{z}^* = [\mathbf{x}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, \boldsymbol{\gamma}^*, \boldsymbol{\eta}^*, \boldsymbol{\psi}^*, \mathbf{s}^*]$, ensuring $0 = K_\rho(\mathbf{z}^*)$. This approach guarantees that the associated quasi-policies form a set of local FSE policies, provided that we anneal the parameter ρ to zero and the sufficient condition in Theorem 4.3 is satisfied. We formalize our method in Algorithm 5.1.

In Algorithm 5.1, we gradually decay the homotopy parameter ρ to zero such that $\lim_{\rho\to 0} \mathbf{z}_\rho$ recovers an FSE solution. For each ρ, at the kth iteration, we first construct the KKT conditions $0 = K_\rho(\mathbf{z})$ along the trajectory $\mathbf{z}_\rho^{(k)}$. We compute the Newton update direction $\Delta\mathbf{z} := -(\nabla K_\rho)^+ \cdot K_\rho(\mathbf{z}_\rho^{(k)})$. Since we aim at finding a solution $\mathbf{z}^*$ to $0 = K_\rho(\mathbf{z}^*)$, a natural choice of merit function is $\|K_\rho(\mathbf{z})\|_2$. Given this choice of the merit function, we perform a line search to determine a step size α and update $\mathbf{z}_\rho^{(k+1)} = \mathbf{z}_\rho^{(k)} + \alpha\Delta\mathbf{z}$ until convergence. The converged solution is denoted as $\mathbf{z}_\rho^*$. Subsequently, we steadily decay ρ and repeat these Newton update steps. We characterize in the following result how the magnitude of the KKT residual value $\|K_\rho(\mathbf{z})\|_2$ influences the convergence rate of Algorithm 5.1 when solving LQ games.

THEOREM 5.1. *Under Assumption* 3.1*, let* $\mathcal{F}_z := \{\mathbf{z} = [\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\psi}, \mathbf{s}] : \boldsymbol{\gamma} \geq 0, \mathbf{s} \geq 0\}$ *be the solution set. We denote by* $\nabla K_\rho(\mathbf{z})$ *and* $\nabla^* K_\rho(\mathbf{z})$ *the Jacobians of the KKT conditions with and without considering quasi-policy gradients, respectively. Suppose that* $\nabla K_\rho(\mathbf{z})$ *is invertible and there exist constants* D *and* C *such that*

$$\|(\nabla K_\rho(\mathbf{z}))^{-1}\|_2 \leq D \qquad \forall i \in \mathbf{I}_1^N,\ \forall \mathbf{z} \in \mathcal{F}_z, \tag{5.22a}$$

$$\|\nabla^* K_\rho(\mathbf{z}) - \nabla^* K_\rho(\tilde{\mathbf{z}})\|_2 \leq C\|\mathbf{z} - \tilde{\mathbf{z}}\|_2 \qquad \forall i \in \mathbf{I}_1^N,\ \forall \mathbf{z}, \tilde{\mathbf{z}} \in \mathcal{F}_z. \tag{5.22b}$$

Let $\hat{\alpha} \in [0,1]$ *be the maximum feasible step size for all* $\mathbf{z} \in \mathcal{F}_\mathbf{z}$*, i.e.,* $\hat{\alpha} := \max\{\alpha \in [0,1] : \mathbf{z}, \mathbf{z} + \alpha\Delta\mathbf{z} \in \mathcal{F}_z\}$*. Moreover, suppose* $\|\nabla^* K_\rho(\mathbf{z}) - \nabla K_\rho(\mathbf{z})\|_2 \leq \delta$ *for all* $\mathbf{z} \in \mathcal{F}_\mathbf{z}$ *and* $D \cdot \delta < 1$*. Then, for all* $\mathbf{z} \in \mathcal{F}_z$*, there exists* $\alpha \in [0, \hat{\alpha}]$ *such that*

1. *if* $\|K_\rho(\mathbf{z})\|_2 > \frac{1-D\delta}{D^2C\hat{\alpha}}$*, then* $\|K_\rho(\mathbf{z} + \alpha\Delta\mathbf{z})\|_2 \leq \|K_\rho(\mathbf{z})\|_2 - \frac{(1-D\delta)^2}{2D^2C}$*;*
2. *if* $\|K_\rho(\mathbf{z})\|_2 \leq \frac{1-D\delta}{D^2C\hat{\alpha}}$*, then* $\|K_\rho(\mathbf{z} + \alpha\Delta\mathbf{z})\|_2 \leq \big(1 - \frac{1}{2}\hat{\alpha}(1 - D\delta)\big) \cdot \|K_\rho(\mathbf{z})\|_2$*, and we have exponential convergence.*

Proof. The proof can be found in Appendix A. ∎

Theorem 5.1 suggests that, under certain conditions, the merit function $\|K_\rho(\mathbf{z})\|_2$ decays to zero exponentially fast, and Algorithm 5.1 converges to a solution satisfying the KKT conditions considering the quasi-policy gradients. The above analysis can be considered as an extension of the classical PDIP convergence proof in [9] to constrained feedback Stackelberg games where we consider feedback interaction constraints $0 = u_t^i - \tilde{\pi}_{t,\rho}^i(x_t, u_t^{1:i-1})$ and the quasi-policy gradients. The condition (5.22a) equates to establishing a lower bound for the smallest nonzero singular value of $\nabla K_\rho(\mathbf{z})$. Practically, this can be achieved by adding a minor cost regularization term to the KKT conditions [11]. Moreover, the constant C in (5.22b) depends on the maximum singular values of the Hessians of costs, the Jacobian of constraints, and linear dynamics, which are all constant matrices in LQ games and can therefore be upper bounded.

Given a $\rho > 0$, a converged solution $\mathbf{z}_\rho^*$ renders $K_\rho(\mathbf{z}_\rho^*) = 0$. Note that the KKT conditions $0 = K_\rho(\mathbf{z}_\rho^*)$ reduce to the one in Theorem 4.2 when ρ decays to zero. As ρ approaches zero, the solution $\mathbf{z}_\rho^*$, when converged, recovers a solution to the KKT conditions in Theorem 4.2. When the sufficient conditions in Theorem 4.3 are also satisfied, the computed solution converges to a local FSE.

6. From LQ games to nonlinear games. In this section, we extend our solution for LQ games to feedback Stackelberg games with nonlinear dynamics. Without loss of generality, each player could have nonquadratic costs. Coupled nonlinear equality and inequality constraints could also exist among players.

6.1. Iteratively approximating nonlinear games via LQ games by aligning their KKT conditions. In this subsection, we introduce a procedure which iteratively approximates the constrained nonlinear games using constrained LQ games, and computes approximate local feedback Stackelberg equilibria for the nonlinear games. These LQ game approximations are designed to ensure that the first-order approximations of their KKT conditions, expressed as $0 = \nabla K_\rho(\mathbf{z}) \cdot \Delta\mathbf{z} + K_\rho(\mathbf{z})$, align with those of the original nonlinear games, specifically considering the inclusion of quasi-policies. Our approach differs from the existing iterative LQ game approximation techniques [19, 22] for FSE policies, which linearize the dynamics and quadraticize only the costs. In contrast, our method linearizes the dynamics but quadraticizes the Lagrangian. This enables us to utilize the convergence results for LQ games, as discussed in the previous section, to analyze the convergence properties of our method in nonlinear games. Consequently, our work provides the first iterative LQ game approximation approach that has provable convergence guarantees for constrained nonlinear feedback Stackelberg games.

In what follows, we introduce local LQ game approximations of the original nonlinear game. Let $\mathbf{z}$ be a solution in the set $\mathcal{F}_z$. We first define the following linear approximation of the dynamics and constraints around $\mathbf{z}$ for all $t \in \mathbf{I}_0^T, i \in \mathbf{I}_1^N$:

(6.1)
$$
\begin{aligned}
&A_t := \nabla_{x_t} f_t(x_t, u_t), && B_t^i := \nabla_{u_t^i} f_t(x_t, u_t), && c_t := f_t(x_t, u_t) - x_{t+1},\\
&H_{x_t}^i := \nabla_{x_t} h_t^i,\ H_{u_t^j}^i := \nabla_{u_t^j} h_t^i, && G_{x_t}^i := \nabla_{x_t} g_t^i,\ G_{u_t^j}^i := \nabla_{u_t^j} g_t^i && \forall j \in \mathbf{I}_1^N,\\
&\bar{h}_t^i := h_t^i(x_t, u_t), && \bar{g}_t^i := g_t^i(x_t, u_t), &&\\
&H_{x_{T+1}}^i := \nabla_{x_{T+1}} h_{T+1}^i, && G_{x_{T+1}}^i := \nabla_{x_{T+1}} g_{T+1}^i, &&\\
&\bar{h}_{T+1}^i := h_{T+1}^i(x_{T+1}), && \bar{g}_{T+1}^i := g_{T+1}^i(x_{T+1}). &&
\end{aligned}
$$

For each $i \in \mathbf{I}_1^N$ and $t \in \mathbf{I}_0^T$, we represent the second-order terms and cost-related terms in the Lagrangian $\mathcal{L}_t^i$ as quadratic costs (5.2), with parameters defined as follows:

$$
\begin{aligned}
Q_t^i &:= \nabla_{xx}^2 \ell_t^i + (\nabla_{xx}^2 f_t)^\top \lambda_t^i - (\nabla_{xx}^2 h_t^i)^\top \mu_t^i - (\nabla_{xx}^2 g_t^i)^\top \gamma_t^i,\\
S_t^i &:= \nabla_{ux}^2 \ell_t^i + (\nabla_{ux}^2 f_t)^\top \lambda_t^i - (\nabla_{ux}^2 h_t^i)^\top \mu_t^i - (\nabla_{ux}^2 g_t^i)^\top \gamma_t^i,\\
R_t^i &:= \nabla_{uu}^2 \ell_t^i + (\nabla_{uu}^2 f_t)^\top \lambda_t^i - (\nabla_{uu}^2 h_t^i)^\top \mu_t^i - (\nabla_{uu}^2 g_t^i)^\top \gamma_t^i,\\
Q_{T+1}^i &:= \nabla_{xx}^2 \ell_{T+1}^i - (\nabla_{xx}^2 h_{T+1}^i)^\top \mu_{T+1}^i - (\nabla_{xx}^2 g_{T+1}^i)^\top \gamma_{T+1}^i,\\
q_t^i &:= \nabla_x \ell_t^i, \qquad r_t^i := \nabla_u \ell_t^i, \qquad q_{T+1}^i := \nabla_x \ell_{T+1}^i.
\end{aligned} \tag{6.2}
$$

We can modify Algorithm 5.1 to address nonlinear games by applying an LQ game approximation around the solution $\mathbf{z}_\rho^{(k)}$ in step 3 of Algorithm 5.1 and formulate the resulting approximate KKT conditions $0 = K_\rho(\mathbf{z}_\rho^{(k)})$ defined with terms in (6.1) and (6.2). Furthermore, this LQ game approximation is reiterated around $\mathbf{z}_\rho^{(k)} + \alpha\Delta\mathbf{z}_\rho$ in step 6, when we evaluate the merit function $\|K_\rho(\mathbf{z}_\rho^{(k)} + \alpha\Delta\mathbf{z}_\rho)\|_2$ during line search.

6.2. Quasi-policy approximation error and exponential convergence analysis in nonlinear games. In the above solution procedure, we approximate the ground truth nonlinear policies of nonlinear games by quasi-policies. However, different from LQ games, the ground truth feedback Stackelberg policies for nonlinear

games could have nonzero high-order policy gradients. Thus, it is worthwhile to characterize the error caused by the quasi-policy gradients. Essentially, there are two error sources. The first type of error is due to the fact that we have neglected high-order policy gradients when evaluating the KKT Jacobian $\nabla K^i_{t,\rho}(\mathbf{z})$, and the second form of error is how these changes propagate into the expression of KKT conditions $0 = K^i_{t,\rho}(\mathbf{z})$ for earlier players and stages. Suppose those two error sources could be upper bounded; then, we can characterize their impact on the policy gradients error in the following proposition.

PROPOSITION 6.1. *Under Assumption* 3.1, *let* $\mathbf{z}$ *and* $\tilde{\mathbf{z}}$ *be two elements in the solution set* $\mathcal{F}_z$. *We denote by* $\{\pi^i_{t,\rho}\}^{T,N}_{t=0,i=1}$ *a set of policies around* $\mathbf{z}$ *and by* $\{\tilde{\pi}^i_{t,\rho}\}^{T,N}_{t=0,i=1}$ *a set of quasi-policies around* $\tilde{\mathbf{z}}$, *respectively. We denote by* $\{K^i_{t,\rho}(\mathbf{z})\}^{T,N}_{t=0,i=1}$ *and* $\{K^{i*}_{t,\rho}(\mathbf{z})\}^{T,N}_{t=0,i=1}$ *the KKT conditions with and without quasi-policies, respectively. Let* $i \le N$ *and* $t \le T$. *Suppose that the Jacobian matrices* $\nabla K^i_{t,\rho}(\tilde{\mathbf{z}})$, $\nabla K^{i*}_{t,\rho}(\tilde{\mathbf{z}})$, *and* $\nabla K^{i*}_{t,\rho}(\mathbf{z})$ *are invertible. Let* $\epsilon_{\mathbf{z},\tilde{\mathbf{z}}} > 0$ *be an upper error bound such that*

$$\begin{aligned}\max\Big\{&\|\nabla K^i_{t,\rho}(\tilde{\mathbf{z}}) - \nabla K^{i*}_{t,\rho}(\tilde{\mathbf{z}})\|_2, \quad \|\nabla K^{i*}_{t,\rho}(\tilde{\mathbf{z}}) - \nabla K^{i*}_{t,\rho}(\mathbf{z})\|_2,\\ &\|K^i_{t,\rho}(\tilde{\mathbf{z}}) - K^i_{t,\rho}(\mathbf{z})\|_2, \qquad \|K^i_{t,\rho}(\mathbf{z}) - K^{i*}_{t,\rho}(\mathbf{z})\|_2\Big\} \le \epsilon_{\mathbf{z},\tilde{\mathbf{z}}}.\end{aligned} \tag{6.3}$$

Then, the error between the quasi-policy gradient and the policy gradient can be bounded as follows:

$$\begin{aligned}\|\nabla\tilde{\pi}^i_{t,\rho}(\tilde{\mathbf{z}}) - \nabla\pi^i_{t,\rho}(\mathbf{z})\|_2 \le \epsilon_{\mathbf{z},\tilde{\mathbf{z}}} \cdot \Big(&2\|\nabla K^{i*}_{t,\rho}(\mathbf{z})^{-1}\|_2\\ &+ \big(\|\nabla K^i_{t,\rho}(\tilde{\mathbf{z}})^{-1}\|_2 + \|\nabla K^{i*}_{t,\rho}(\mathbf{z})^{-1}\|_2\big) \cdot \|\nabla K^{i*}_{t,\rho}(\tilde{\mathbf{z}})^{-1}\|_2 \|K^i_{t,\rho}(\tilde{\mathbf{z}})\|_2\Big).\end{aligned} \tag{6.4}$$

Proof. The proof can be found in Appendix A. ∎

Proposition 6.1 suggests that the error introduced by the quasi-policy gradients is proportional to $\epsilon_{\mathbf{z},\tilde{\mathbf{z}}}$, as described in (6.4). However, it is challenging to obtain an analytical bound $\epsilon_{\mathbf{z},\tilde{\mathbf{z}}}$ because the evaluation of $K^{i*}_{t,\rho}(\mathbf{z})$ and $\nabla K^{i*}_{t,\rho}(\mathbf{z})$ requires computing the high-order policy gradients. The above analysis only provides a partial analysis for the policy gradient error introduced by the quasi-policy gradients. In principle, it is possible that the quasi-policy gradients could lead to a different feedback Stackelberg policy from the ground truth feedback Stackelberg policy. However, it is intractable to compute high-order policy gradients when we have a long-horizon game. In general, the quasi-policy is a local linear approximation to the ground truth nonlinear feedback Stackelberg policy, and when a state perturbation occurs at time t, such policies are only approximately optimal for the resulting subgame. We believe that the local feedback Stackelberg quasi-policy is the closest computationally tractable approximation possible when we consider the first-order policy approximation techniques for long-horizon feedback Stackelberg games.

Furthermore, we can leverage the sufficient condition of the local FSE and the convergence analysis in Theorem 5.1 to show that we will converge to a local FSE of nonlinear games under certain conditions on the iterative LQ game approximations.

THEOREM 6.2 (exponential convergence in nonlinear games). *Suppose that there exist constants* $(D, C, \delta, \hat{\alpha})$, *as defined in Theorem* 5.1, *such that at each iteration* k *of Algorithm* 5.1, *the approximate LQ game defined in* (6.1) *and* (6.2) *satisfies the conditions of Theorem* 5.1. *Then, for each* $\rho > 0$ *and a sufficiently large* k, $\mathbf{z}^{(k)}_\rho$ *converges exponentially fast to a solution* $\mathbf{z}^*_\rho$, *which renders* $\|K_\rho(\mathbf{z}^*_\rho)\|_2 = 0$. *Moreover, if the*

limit $\mathbf{z}^* := \lim_{\rho\to 0}\mathbf{z}^*_\rho$ *exists and Theorem* 4.3*, which provides a sufficient condition for local FSE trajectories, holds true at* $\mathbf{z}^*_\rho$ *for all* $\rho > 0$*, then the converged solution* $\mathbf{z}^*$ *recovers a local FSE trajectory.*

Proof. The proof can be found in Appendix A. ∎

7. Experiments. In this section, we consider a two-player feedback Stackelberg game modeling highway driving,[1] where two highway lanes merge into one and the planning horizon $T = 20$. We associate with each player a 4-dimensional state vector $x^i_t = [p^i_{x,t}, p^i_{y,t}, v^i_t, \theta^i_t]$, where $(p^i_{x,t}, p^i_{y,t})$ represents the (x, y) coordinate, v^i_t denotes the velocity, and θ^i_t encodes the heading angle of player i at time t. The joint state vector of the two players is denoted as $x_t = [x^1_t, x^2_t]$. Both players have nonlinear unicycle dynamics: $\forall t \in \mathbf{I}_0^T$, $\forall i \in \{1, 2\}$,

$$\begin{aligned} &p^i_{x,t+1} = p^i_{x,t} + \Delta t \cdot v^i_t \sin(\theta^i_t), \quad p^i_{y,t+1} = p^i_{y,t} + \Delta t \cdot v^i_t \cos(\theta^i_t), \\ &v^i_{t+1} = v^i_t + \Delta t \cdot a^i_t, \qquad\qquad\quad \theta^i_{t+1} = \theta^i_t + \Delta t \cdot \omega^i_t. \end{aligned} \tag{7.1}$$

We consider the cost functions, for all $t \in \mathbf{I}_0^T$,

$$\ell^1_t(x_t, u_t) = 10(p^1_{x,t} - 0.4)^2 + 6(v^1_t - v^2_t)^2 + 2\|u^1_t\|_2^2, \ell^2_t(x_t, u_t) = \|\theta^2_t\|_2^4 + 2\|u^2_t\|_2^2, \tag{7.2}$$

and the terminal costs $\ell^1_{T+1}(x_{T+1}) = 10(p^1_{x,T+1} - 0.4)^2 + 6(v^1_t - v^2_t)^2$ and $\ell^2_{T+1}(x_{T+1}) = \|\theta^2_{T+1}\|_2^4$. Note that we include a fourth-order cost term in player 2's cost at each stage to model its preference of small heading angle. We consider the following (nonconvex) constraints encoding collision avoidance, driving on the road, and control limits:

$$\begin{aligned} &\sqrt{\|p^1_{x,t} - p^2_{x,t}\|_2^2 + \|p^1_{y,t} - p^2_{y,t}\|_2^2} - d_{\text{safe}} \geq 0, && t \in \mathbf{I}_0^{T+1}, \\ &p^i_{x,t} - p_l \geq 0, \qquad p_r(p^i_{y,t}, p^i_{x,t}) \geq 0, \qquad \|u_t\|_\infty \leq u_{\max}, && t \in \mathbf{I}_0^{T+1},\ i \in \{1, 2\}, \end{aligned} \tag{7.3}$$

where we define $p_l \in \mathbb{R}$ to be the left road boundary and denote by $p_r(p^i_{x,t}, p^i_{y,t})$ the distance between player i and the right road boundary curve. We also consider the following equality constraints at the terminal time:

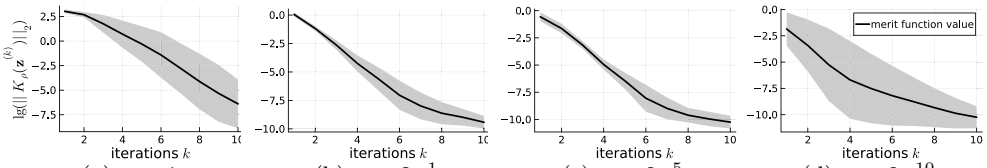
$$v^1_{T+1} - v^2_{T+1} = 0, \quad \theta^1_{T+1} = 0, \tag{7.4}$$

where the two players aim to reach a consensus on their speeds, with player 1 maintaining its heading angle pointing forward.

The nominal initial states of two players are specified as $x^1_0 = [0.9, 1.2, 3.5, 0.0]$ and $x^2_0 = [0.5, 0.6, 3.8, 0.0]$, respectively. We randomly sample 10 initial states around $x_0 = [x^1_0, x^2_0]$ under a uniform distribution within the range of -0.1 to 0.1. From each sampled $\hat{x}_0$, we obtain an initial state trajectory $\mathbf{x}^{(0)}$ by simulating the nonlinear dynamics (7.1) with the initial controls $\mathbf{u}^{(0)} = \mathbf{0}$. Set the initial slack variables for the inequality constraints as $\mathbf{s}^{(0)} = \mathbf{1}$, along with the corresponding Lagrange multipliers $\boldsymbol{\gamma}^{(0)} = \mathbf{1}$. We set all other Lagrange multipliers $\{\boldsymbol{\lambda}^{(0)}, \boldsymbol{\mu}^{(0)}, \boldsymbol{\eta}^{(0)}, \boldsymbol{\psi}^{(0)}\}$ to zeros. Consequently, we have constructed an initial solution $\mathbf{z}^{(0)} = [\mathbf{x}^{(0)}, \mathbf{u}^{(0)}, \boldsymbol{\lambda}^{(0)}, \boldsymbol{\gamma}^{(0)}\boldsymbol{\mu}^{(0)}, \boldsymbol{\eta}^{(0)}, \boldsymbol{\psi}^{(0)}, \mathbf{s}^{(0)}]$. We repeat this initialization trajectory defining process for different sampled $\hat{x}_0$.

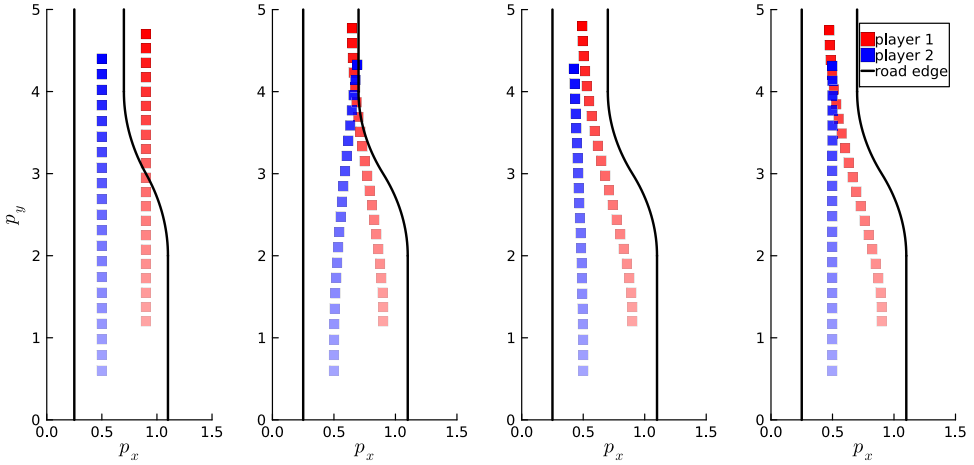
For each sampled initial state $\hat{x}_0$, we employ Algorithm 5.1 with iterative LQ game approximations to compute a local FSE trajectory. The convergence of our method under different sampled $\hat{x}_0$ is depicted in Figure 1. For each ρ, the merit function value decreases as the iterations continue. Furthermore, since the cost functions are

[1]The code is available at https://github.com/jamesjingqili/FeedbackStackelbergGames.jl.git.



(a) $\rho = 1$. (b) $\rho = 2^{-1}$. (c) $\rho = 2^{-5}$. (d) $\rho = 2^{-10}$.

FIG. 1. *Convergence of Algorithm* 5.1 *with iterative LQ game approximations under different values of the homotopy parameter* ρ *from* 10 *sampled initial states. The solid curve and the shaded area denote the mean and the standard deviation of the logarithm of the merit function values, respectively. By gradually annealing* ρ *to zero, the solution converges to a local FSE trajectory. Moreover, under each* ρ*, the plots above empirically support the linear convergence described in Theorem* 6.2.



(a) $\rho = 1$, $k = 0$. (b) $\rho = 1$, $k = 3$. (c) $\rho = 1$, $k = 6$. (d) $\rho = 2^{-10}$, $k = 10$.

FIG. 2. *Tolerance of an infeasible trajectory initialization and the converged trajectories of two players. In Figure* 2(a)*, we plot the initial state trajectories of two players, where player* 1*'s trajectory is infeasible because it violates the road boundary constraint. When* $\rho = 1$*, we plot the state trajectories in the third and sixth iterations in Figures* 2(b) *and* 2(c)*, respectively. They become feasible at the sixth iteration. In Figure* 2(d)*, we plot the converged solution, with* $\rho = 2^{-10}$.

strongly convex with respect to each player's controls, Theorem 4.3 ensures that our converged solution constitutes a local FSE trajectory. Moreover, we show our method can tolerate infeasible initialization in Figure 2, where the right road boundary constraint is initially violated by initialization $\mathbf{z}^{(0)}$, and as the algorithm progresses, subsequent iterates $\mathbf{z}^{(k)}$ become feasible.

8. Conclusions. In this paper, we considered general-sum feedback Stackelberg dynamic games with coupled constraints among N players. We proposed a primal-dual interior point method to compute an approximate feedback Stackelberg equilibrium and the associated policies for all players. To the best of the authors' knowledge, this represents the first attempt to compute approximate local feedback Stackelberg equilibria in both linear quadratic games and nonlinear games under general coupled equality and inequality constraints, within continuous state and action spaces. We theoretically characterized the approximation error and the exponential convergence of our algorithm. Numerical experiments suggest that the proposed algorithm can tolerate infeasible initializations and efficiently converge to a feasible equilibrium so-

lution. Future research should investigate the potential benefits of higher-order policy gradient approximations. Additionally, extending our approach to solve other types of equilibria in dynamic games is also a promising direction for future research.

Appendix A. Supplementary results.

Proof of Theorem 4.1. At the terminal time $t = T$, for ease of notation, we define $x_T = \bar{x}_T$ and $u_T^{1:i-1} = \bar{u}_T^{1:i-1}$. We observe that, for each player $i \in \mathbf{I}_1^N$, (4.1) can be rewritten as

$$\begin{aligned}\tilde{u}_T^i \in \arg\min_{u_T^i} \min_{\substack{u_T^{i+1:N} \\ x_{T+1}}} & \left\{ \ell_T^i(x_T, u_T) + V_{T+1}^i(x_{T+1}) \right\} \\ \text{s.t. } & 0 = u_T^j - \pi_T^j(x_T, u_T^{1:j-1}),\ 0 = x_{T+1} - f_T(x_T, u_T), \quad j \in \mathbf{I}_{i+1}^N, \\ & 0 = h_T^i(x_T, u_T),\ 0 \le g_T^i(x_T, u_T), \\ & 0 = h_{T+1}^i(x_{T+1}),\ 0 \le g_{T+1}^i(x_{T+1}),\end{aligned}$$

which implies $\tilde{u}_T^i \in \arg_{u_T^i} \min_{u_T^i} Z_T^i(\bar{x}_T, \bar{u}_T^{1:i-1}, u_T^i)$. Moreover, for all $t \in \mathbf{I}_0^{T-1}$ and $i \in \mathbf{I}_1^N$, for the ease of notation, we assume $x_t = \bar{x}_t$ and $u_t^{1:i-1} = \bar{u}_t^{1:i-1}$. We observe

$$\begin{aligned}\tilde{u}_t^i \in \arg\min_{u_t^i} \min_{\substack{u_t^{i+1:N} \\ u_{t+1:T}^{1:N} \\ x_{t+1:T+1}}} & \left\{ \sum_{\tau=t}^{T} \ell_\tau^i(x_\tau, u_\tau) + \ell_{T+1}^i(x_{T+1}) \right\} \\ \text{s.t. } & 0 = u_t^j - \pi_t^j(x_t, u_t^{1:j-1}), && j \in \mathbf{I}_{i+1}^N, \\ & 0 = x_{\tau+1} - f_\tau(x_\tau, u_\tau), && \tau \in \mathbf{I}_t^T, \\ & 0 = u_\tau^j - \pi_\tau^j(x_\tau, u_\tau^{1:j-1}), && \tau \in \mathbf{I}_{t+1}^T,\ j \in \mathbf{I}_1^N \setminus \{i\}, \\ & 0 = h_\tau^i(x_\tau, u_\tau),\ 0 \le g_\tau^i(x_\tau, u_\tau), && \tau \in \mathbf{I}_t^T, \\ & 0 = h_{T+1}^i(x_{T+1}),\ 0 \le g_{T+1}^i(x_{T+1}).\end{aligned}$$

The above can be further rewritten as

$$\begin{aligned}\tilde{u}_t^i \in \arg\min_{u_t^i} \min_{\substack{u_t^{i+1:N} \\ x_{t+1}}} & \left\{ \ell_t^i(x_t, u_t) + V_{t+1}^i(x_{t+1}) \right\} \\ \text{s.t. } & 0 = u_t^j - \pi_t^j(x_t, u_t^{1:j-1}),\ 0 = x_{t+1} - f_t(x_t, u_t), \quad j \in \mathbf{I}_{i+1}^N, \\ & 0 = h_t^i(x_t, u_t),\ 0 \le g_t^i(x_t, u_t).\end{aligned}$$

It follows that $\tilde{u}_t^i \in \arg_{u_t^i} \min_{u_t^i} Z_t^i(\bar{x}_t, \bar{u}_t^{1:i-1}, u_t^i)$. Therefore, the set of strategies $\{\pi_t^i\}_{t=0,i=1}^{T,N}$ constitutes a set of local feedback Stackelberg policies. ∎

Proof of Theorem 4.2. For a time $t \in \mathbf{I}_0^T$ and player $i \in \mathbf{I}_1^N$, we set the gradient of $\mathcal{L}_t^i$ with respect to u_t^i and x_t to be zero. This constitutes the first two rows of (4.5). In addition, a player $i < N$ considers the feedback interaction constraints $0 = u_t^{j*} - \pi_t^j(x_t^*, u_1^{1:j-1*})$ for $j \in \mathbf{I}_{i+1}^N$. This constraint is implicitly ensured when we enforce player j's KKT conditions in player i's KKT conditions. Thus, we only need to ensure the gradient $\nabla_{u_t^j} \mathcal{L}_t^i$ to be zero when synthesizing player i's KKT conditions. This corresponds to the third row of (4.5). Moreover, at a time $t < T$, each player $i \in \mathbf{I}_1^N$ needs to account for the feedback reaction from other players in future steps. Again this constraint is implicitly ensured when we define player j's KKT conditions. We only need to additionally set the gradient of $\mathcal{L}_t^i$ with respect to u_τ^j to be zero,

where $\tau \in \mathbf{I}_{t+1}^T$ and $j \in \mathbf{I}_1^N \setminus \{i\}$. These correspond to the fourth row of (4.5). Finally, we include the dynamics constraints, equality and inequality constraints, and complementary slackness conditions in the last five rows of (4.5). ∎

Proof of Theorem 4.3. We can check that the feasible set for the equality constraints of (4.6) is a superset of the critical cone of the problem (4.5). By Theorem 12.6 in [37], the solution $(\mathbf{x}^*, \mathbf{u}^*)$ constitutes a local FSE trajectory. ∎

Proof of Theorem 5.1. By the fundamental theorem of calculus, we have $K_\rho(\mathbf{z} + \alpha\Delta\mathbf{z}) = K_\rho(\mathbf{z}) + \int_0^1 \nabla^* K_\rho(\mathbf{z} + \tau\alpha\Delta\mathbf{z})\alpha\Delta\mathbf{z}d\tau$, and we have

(A.1)
$$\begin{aligned}\|K_\rho(\mathbf{z}+\alpha\Delta\mathbf{z})\|_2 &= \left\|K_\rho(\mathbf{z}) + \int_0^1 \nabla^* K_\rho(\mathbf{z}+\tau\alpha\Delta\mathbf{z})\alpha\Delta\mathbf{z}d\tau\right\|_2 \\ &\le \|K_\rho(\mathbf{z}) + \alpha\nabla^* K_\rho(\mathbf{z})\Delta\mathbf{z}\|_2 + \left\|\int_0^1 (\nabla^* K_\rho(\mathbf{z}+\tau\alpha\Delta\mathbf{z}) - \nabla^* K_\rho(\mathbf{z}))\alpha\Delta\mathbf{z}d\tau\right\|_2.\end{aligned}$$

Substituting $\Delta\mathbf{z}$ into $\|K_\rho(\mathbf{z}) + \alpha\nabla^* K_\rho(\mathbf{z})\Delta\mathbf{z}\|_2$, we have

(A.2)
$$\begin{aligned}\|K_\rho(\mathbf{z}) + \alpha\nabla^* K_\rho(\mathbf{z})\Delta\mathbf{z}\|_2 &= \|K_\rho(\mathbf{z}) - \alpha\nabla^* K_\rho(\mathbf{z})(\nabla K_\rho(\mathbf{z}))^{-1}K_\rho(\mathbf{z})\|_2 \\ &\le (1-\alpha)\|K_\rho(\mathbf{z})\|_2 + \alpha\|\nabla^* K_\rho(\mathbf{z}) - \nabla K_\rho(\mathbf{z})\|_2\|(\nabla K_\rho(\mathbf{z}))^{-1}\|_2\|K_\rho(\mathbf{z})\|_2 \\ &\le (1-\alpha)\|K_\rho(\mathbf{z})\|_2 + \alpha\delta D\|K_\rho(\mathbf{z})\|_2 = (1-\alpha(1-\delta D))\|K_\rho(\mathbf{z})\|_2.\end{aligned}$$

Combining (A.2) and (A.1), we have

$$\begin{aligned}&\|K_\rho(\mathbf{z}+\alpha\Delta\mathbf{z})\|_2 \\ &\le (1-\alpha(1-\delta D))\|K_\rho(\mathbf{z})\|_2 + \|\alpha\Delta\mathbf{z}\|_2\left\|\int_0^1 \|\nabla^* K_\rho(\mathbf{z}+\tau\alpha\Delta\mathbf{z}) - \nabla^* K_\rho(\mathbf{z})\|d\tau\right\|_2 \\ &\le (1-\alpha(1-\delta D))\|K_\rho(\mathbf{z})\|_2 + \frac{1}{2}\alpha^2 D^2 C\|K_\rho(\mathbf{z})\|_2^2,\end{aligned}$$

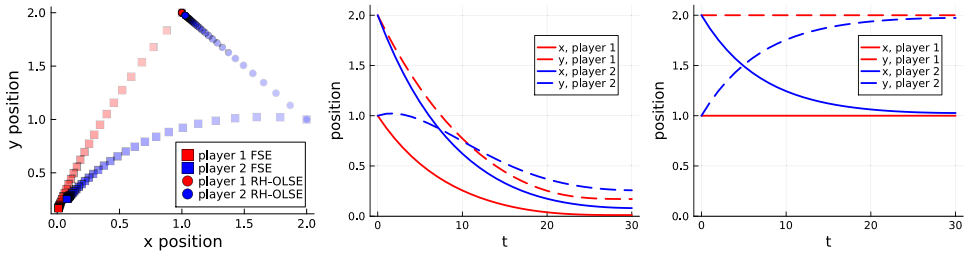
where the right-hand side is minimized when $\alpha^* = \frac{1-D\delta}{D^2C\|K_\rho(\mathbf{z})\|_2}$. Suppose $\|K_\rho(\mathbf{z})\|_2 > \frac{1-D\delta}{D^2C\hat{\alpha}}$; then $\hat{\alpha} > \frac{1-D\delta}{D^2C\|K_\rho(\mathbf{z})\|_2}$ and we have $\|K_\rho(\mathbf{z}+\alpha^*\Delta\mathbf{z})\|_2 \le \|K_\rho(\mathbf{z})\|_2 - \frac{(1-D\delta)^2}{2D^2C}$.

For the case $\|K_\rho(\mathbf{z})\|_2 \le \frac{1-D\delta}{D^2C\hat{\alpha}}$, let $\alpha := \hat{\alpha}$. By $\hat{\alpha}D^2C\|K_\rho(\mathbf{z})\|_2 \le 1 - D\delta$, we have $\|K_\rho(\mathbf{z}+\hat{\alpha}\Delta\mathbf{z})\|_2 \le (1 - \frac{1}{2}\hat{\alpha}(1-D\delta))\|K_\rho(\mathbf{z})\|_2$. ∎

Proof of Proposition 6.1. By definition, we have

$$\begin{aligned}&\|\nabla\tilde{\pi}_{t,\rho}^i(\tilde{\mathbf{z}}) - \nabla\pi_{t,\rho}^i(\mathbf{z})\|_2 = \|\nabla K_{t,\rho}^i(\tilde{\mathbf{z}})^{-1}K_{t,\rho}^i(\tilde{\mathbf{z}}) - \nabla K_{t,\rho}^{i*}(\mathbf{z})^{-1}K_{t,\rho}^{i*}(\mathbf{z})\|_2 \\ &= \|\nabla K_{t,\rho}^i(\tilde{\mathbf{z}})^{-1}K_{t,\rho}^i(\tilde{\mathbf{z}}) - \nabla K_{t,\rho}^{i*}(\mathbf{z})^{-1}K_{t,\rho}^{i*}(\mathbf{z}) \\ &\quad + \nabla K_{t,\rho}^{i*}(\tilde{\mathbf{z}})^{-1}K_{t,\rho}^i(\tilde{\mathbf{z}}) - \nabla K_{t,\rho}^{i*}(\tilde{\mathbf{z}})^{-1}K_{t,\rho}^i(\tilde{\mathbf{z}}) \\ &\quad + \nabla K_{t,\rho}^{i*}(\mathbf{z})^{-1}K_{t,\rho}^i(\tilde{\mathbf{z}}) - \nabla K_{t,\rho}^{i*}(\mathbf{z})^{-1}K_{t,\rho}^i(\tilde{\mathbf{z}}) \\ &\quad + \nabla K_{t,\rho}^{i*}(\mathbf{z})^{-1}K_{t,\rho}^i(\mathbf{z}) - \nabla K_{t,\rho}^{i*}(\mathbf{z})^{-1}K_{t,\rho}^i(\mathbf{z})\|_2 \\ &\le \Big(\|\nabla K_{t,\rho}^i(\tilde{\mathbf{z}})^{-1} - \nabla K_{t,\rho}^{i*}(\tilde{\mathbf{z}})^{-1}\|_2 + \|\nabla K_{t,\rho}^{i*}(\tilde{\mathbf{z}})^{-1} - \nabla K_{t,\rho}^{i*}(\mathbf{z})^{-1}\|_2\Big)\|K_{t,\rho}^i(\tilde{\mathbf{z}})\|_2 \\ &\quad + \Big(\|K_{t,\rho}^i(\tilde{\mathbf{z}}) - K_{t,\rho}^i(\mathbf{z})\|_2 + \|K_{t,\rho}^i(\mathbf{z}) - K_{t,\rho}^{i*}(\mathbf{z})\|_2\Big)\|\nabla K_{t,\rho}^{i*}(\mathbf{z})^{-1}\|_2 \\ &\le \epsilon_{\mathbf{z},\tilde{\mathbf{z}}}\Big(\|\nabla K_{t,\rho}^i(\tilde{\mathbf{z}})^{-1}\|_2 + \|\nabla K_{t,\rho}^{i*}(\mathbf{z})^{-1}\|_2\Big)\|\nabla K_{t,\rho}^{i*}(\tilde{\mathbf{z}})^{-1}\|_2\|K_{t,\rho}^i(\tilde{\mathbf{z}})\|_2 \\ &\quad + 2\epsilon_{\mathbf{z},\tilde{\mathbf{z}}}\|\nabla K_{t,\rho}^{i*}(\mathbf{z})^{-1}\|_2,\end{aligned}$$

where the last line follows by applying Lemma A.1. ∎



(a) Trajectories. (b) Feedback Stackelberg policy. (c) Receding-horizon open-loop Stackelberg policy.

FIG. 3. *The trajectories under the receding-horizon open-loop Stackelberg equilibrium (RH-OLSE) policy and those under the FSE policy are quite different, regardless of the initial conditions. For example, in the above case, under the FSE policy, player* 1 *first moves toward the origin and then player* 2 *follows. However, under the RH-OLSE policy, player* 1 *always stays at its initial position, waiting for player* 2 *to approach.*

LEMMA A.1. *Let K and $\tilde{K}$ be two invertible matrices. Suppose $\|K - \tilde{K}\|_2 \leq \epsilon$; then we have $\|K^{-1} - \tilde{K}^{-1}\|_2 \leq \epsilon \|K^{-1}\|_2 \cdot \|\tilde{K}^{-1}\|_2$.*

Proof of Lemma A.1. Define $\bar{K} := K - \tilde{K}$. Applying the Woodbury matrix equality, we have $\tilde{K}^{-1} = K^{-1} + K^{-1} \cdot \bar{K} \cdot \tilde{K}^{-1}$, and this implies $\|\tilde{K}^{-1} - K^{-1}\| \leq \epsilon \|K^{-1}\| \cdot \|\tilde{K}^{-1}\|$. ∎

Proof of Theorem 6.2. Observe that the first-order approximation of the KKT conditions for the local LQ game approximations coincides with the one for nonlinear games. By Theorem 5.1, for each $\rho > 0$, $\lim_{k\to\infty} \|K_\rho(\mathbf{z}_\rho^{(k)})\|_2 = 0$, and we have exponential convergence when $k \geq \|K_\rho(\mathbf{z}_\rho^{(0)})\|_2 / (\frac{1-D\delta}{D^2 C\hat{\alpha}})$. Moreover, by Theorem 4.3, the solution $\lim_{\rho\to 0} \mathbf{z}_\rho^*$ recovers a local FSE trajectory. ∎

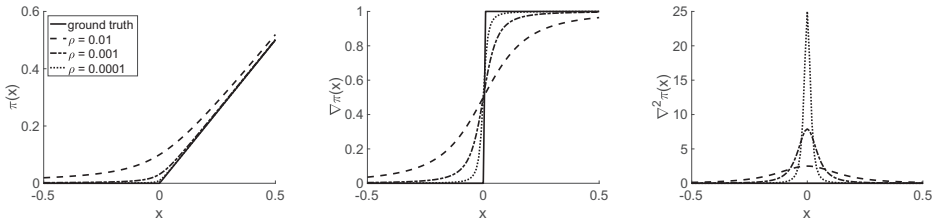
A.1. Comparing the FNE with the FSE. Consider a two-player lane exchanging problem[2] with linear double integrator dynamics. Let $d_{\text{safe}}(x_t) := \frac{1}{2}((p_{x,t}^1 - p_{x,t}^2)^2 + (p_{y,t}^1 - p_{y,t}^2)^2)$ and $d_{\text{target}}(x_t) := (p_{x,t}^1 - 1)^2 + (p_{y,t}^1 - 10)^2 + (p_{x,t}^2 + 1)^2 + (p_{y,t}^2 - 10)^2$. Consider costs $\ell_t^1(x_t, u_t) = d_{\text{target}}(x_t) - d_{\text{safe}}(x_t) + (v_{x,t}^1)^2 + (v_{y,t}^1 - 1)^2 + 4\|u_t^1\|_2^2$ and $\ell_t^2(x_t, u_t) = d_{\text{target}}(x_t) - d_{\text{safe}}(x_t) + (v_{x,t}^2)^2 + (v_{y,t}^2 - 1)^2 + 4\|u_t^2\|_2^2$. Figure 5 suggests that the FSE is a more appropriate equilibrium concept than the FNE when decision hierarchy exists.

A.2. A counterexample that the receding-horizon open-loop Stackelberg equilibrium fails to approximate the FSE well. We consider Example 1 from [28]. We show in Figure 3(a) that the receding-horizon open-loop Stackelberg policy could lead to a trajectory quite different from the one under FSE. Therefore, it is crucial to study the computation of FSE.

A.3. The decay of high-order policy gradients when we apply PDIP to solve constrained LQ games. We validate the quasi-policy assumption in LQ games in Proposition A.2 and include a simplified example in Figure 4.

PROPOSITION A.2. *Under the same assumptions of Theorem* 5.1, *let $\rho > 0$ and denote by $\mathbf{z}_\rho^*$ a converged solution to an LQ game under Algorithm* 5.1 *with high-order policy gradients being considered. Let $\{\pi_{t,\rho}^i\}_{t=0,i=1}^{T,N}$ be the converged policies. Suppose that $\lim_{\rho\to 0} \mathbf{z}_\rho^*$ exists and we denote it by $\mathbf{z}^*$. Moreover, suppose that the ground truth*

[2]The code is available at https://github.com/jamesjingqili/FeedbackStackelbergGames.jl.git.



(a) Zeroth order. (b) First order. (c) Second order.

FIG. 4. *Visualization of the policy gradients of a constrained single-stage linear quadratic regulator problem under different values of ρ. The cost is given by $(u_0 - x_0)^2$. The dynamics is defined as $x_1 = x_0 + u_0$. We consider a constraint $u_0 \geq 0$. The ground truth piecewise linear policy is not differentiable at $x = 0$. As $\rho \to 0$, the policy obtained from PDIP and its first-order gradient closely approximate the ground truth policy and its first-order gradient for all nonzero x. As shown in Figure* 4(c)*, the high-order gradient of the PDIP policy decays to zero as $\rho \to 0$ for all nonzero x.*

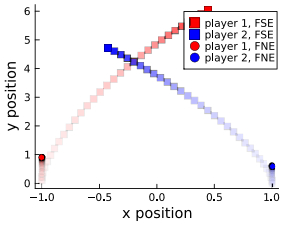


FIG. 5. *Players correctly exchange lanes under the FSE policy but fail to do so under the FNE policy due to safety concern.*

FSE policies $\{\pi_t^i\}_{t=0,i=1}^{T,N}$ *are differentiable at* $(\mathbf{x}^*, \mathbf{u}^*)$*. Then,* $\lim_{\rho\to 0} \|\nabla\pi_{t,\rho}^i - \nabla\pi_t^i\|_2 = 0$ *and* $\lim_{\rho\to 0} \|\nabla^j\pi_{t,\rho}^i\|_2 = 0 \ \ \forall i \in \mathbf{I}_1^N, t \in \mathbf{I}_0^T,\ j \geq 2$.

Proof. At time $t = T$, there is no policy gradient term in the Nth player's KKT conditions. Recall that $\nabla\pi_{T,\rho}^N = -[(\nabla K_{T,\rho}^N)^{-1}]_{u_T^N} \nabla_{[x_T, u_T^{1:N-1}]} K_{T,\rho}^N$ and $\nabla\pi_T^N = -[(\nabla K_T^N)^{-1}]_{u_T^N} \nabla_{[x_T, u_T^{1:N-1}]} K_T^N$. Since $\lim_{\rho\to 0} \|K_{T,\rho}^N(\mathbf{z}_{T,\rho}^{N*}) - K_T^N(\mathbf{z}_T^{N*})\|_2 = 0$, we have pointwise convergence $\lim_{\rho\to 0} \|\nabla\pi_{T,\rho}^N - \nabla\pi_T^N\|_2 = 0$ almost everywhere. We characterize those high-order quasi-policy gradients of $\pi_{T,\rho}^N$ as follows. We denote the map from $\mathbf{z}_{T,\rho}^{N*}$ to the jth-order gradient of $\pi_{T,\rho}^N$ by an operator $\mathcal{A}_T^{N,j} : \mathbf{z}_{T,\rho}^{N*} \to \nabla^j\pi_{T,\rho}^N$. Observe that $[(\nabla K_{T,\rho}^N)^{-1}]_{u_T^N}$ can be considered as the concatenation of a matrix inverse operator $\mathcal{M} : X \in \mathbb{R}^{n\times n} \to X^{-1} \in \mathbb{R}^{n\times n}$ and a linear operator $\hat{\mathcal{M}} : \mathbf{z}_{T,\rho}^{N*} \to \nabla K_{T,\rho}^N$. Note that the matrix inverse is an infinitely differentiable operator when X is invertible and $\nabla_{[x_T, u_T^{1:N-1}]} K_{T,\rho}^N$ is a constant matrix. Thus, by the chain rule [41], $\pi_{T,\rho}^N$ is infinitely differentiable, which also implies that $\nabla^j\pi_{T,\rho}^N$ is continuous $\forall j \geq 1$.

Since $\nabla K_{T,\rho}^N(\mathbf{z}_{T,\rho}^{N*})$ is invertible at $\mathbf{z}_{T,\rho}^{N*}$ and $\mathcal{A}_T^{N,j}$ $\forall j \geq 1$ is a continuous operator, there exists a compact set $\mathcal{S}$ containing $\mathbf{z}_{T,\rho}^{N*}$ such that $\nabla K_{T,\rho}^N(\mathbf{z}_T^N)$ is invertible for all $\mathbf{z}_T^N \in \mathcal{S}$. By the compactness of $\mathcal{S}$ and the continuity of $\mathcal{A}_T^{N,j}$, we have that $\mathcal{A}_T^{N,j}$ is a uniformly continuous operator on $\mathcal{S}$. By Theorem 2 in [3], a uniformly continuous operator preserves the pointwise convergence. Thus, $\lim_{\rho\to 0} \|\nabla^j\pi_{T,\rho}^N - \nabla^j\pi_T^N\|_2 = 0$. Since the ground truth policy π_T^N is piecewise linear and the high-order gradients of π_T^N vanish, we have $\lim_{\rho\to 0} \|\nabla^j\pi_{T,\rho}^N\|_2 = 0$ $\forall j > 1$.

Subsequently, for player $i = N - 1$, since $\lim_{\rho\to 0} \|\nabla\pi_{T,\rho}^N - \nabla\pi_T^N\|_2 = 0$, we have $\lim_{\rho\to 0} \|\nabla K_T^i(\mathbf{z}_{T,\rho}^{i*}) - \nabla K_T^i(\mathbf{z}_T^{i*})\|_2 = 0$, which implies $\lim_{\rho\to 0} \|\nabla\pi_{T,\rho}^i - \nabla\pi_T^i\|_2 = 0$.

A similar reasoning as above yields that $\lim_{\rho\to\infty} \|\nabla^j \pi^i_{T,\rho} - \nabla^j \pi^i_T\|_2 = 0 \ \forall j > 1$. Moreover, we can show that for all players $i < N-1$, $\lim_{\rho\to 0} \|\nabla^j \pi^i_{T,\rho} - \nabla^j \pi^i_T\|_2 = 0$ $\forall j \geq 1$. We continue this backward induction proof of $\lim_{\rho\to 0} \|\nabla^j \pi^i_{t,\rho} - \nabla^j \pi^i_t\|_2 = 0$ $\forall j \geq 1$ for prior stages backward in player decision order until $t = 0$ and $i = 1$. ∎

Appendix B. KKT conditions for two-player LQ games. The KKT conditions $0 = K^2_{T,\rho}(\mathbf{z}^2_T)$ of player 2 at time T are

$$\begin{cases} 0 = \Sigma^2_{j=1} R^{2,2,j}_T u^j_T + S^{2,2}_T x_T + r^{2,2}_T + B^{2\top}_T \lambda^2_T - G^{2\top}_{u^2_T} \gamma^2_T - H^{2\top}_{u^2_T} \mu^2_T, \\ 0 = Q^2_{T+1} x_{T+1} + q^2_{T+1} - \lambda^2_T - G^{2\top}_{x_{T+1}} \gamma^2_{T+1} - H^{2\top}_{x_{T+1}} \mu^2_{T+1}, \\ 0 = x_{T+1} - A_T x_T - B^1_T u^1_T - B^2_T u^2_T - c_T, \\ 0 = H^2_{u^2_T} u^2_T + H^2_{x_T} x_T + H^2_{u^1_T} u^1_T + \bar{h}^2_T, \\ 0 = H^2_{x_{T+1}} x_{T+1} + \bar{h}^2_{T+1}, \\ 0 = \gamma^2_{T:T+1} \odot s^2_{T:T+1} - \rho \mathbf{1}, \\ 0 = G^2_{u^2_T} u^2_T + G^2_{x_T} x_T + G^2_{u^1_T} u^1_T + \bar{g}^2_T - s^2_T, \\ 0 = G^2_{x_{T+1}} x_{T+1} + \bar{g}^2_{T+1} - s^2_{T+1}. \end{cases}$$

We construct the KKT conditions $0 = K^1_{T,\rho}(\mathbf{z}^1_T)$ of player 1 at time T:

$$\begin{cases} 0 = \Sigma^2_{j=1} R^{1,1,j}_T u^j_T + S^{1,1}_T x_T + r^{1,1}_T + B^{1\top}_T \lambda^1_T - G^{1\top}_{u^1_T} \gamma^1_T - H^{1\top}_{u^1_T} \mu^1_T + (\nabla_{u^1_T} \pi^2_{T,\rho})^\top \psi^1_T, \\ 0 = Q^1_{T+1} x_{T+1} + q^1_{T+1} - \lambda^1_T - G^{1\top}_{x_{T+1}} \gamma^1_{T+1} - H^{1\top}_{x_{T+1}} \mu^1_{T+1}, \\ 0 = \Sigma^2_{j=1} R^{1,2,j}_T u^j_T + S^{1,2}_T x_T + r^{1,2}_T + B^{2\top}_T \lambda^1_T - G^{1\top}_{u^2_T} \gamma^1_T - H^{1\top}_{u^2_T} \mu^1_T - \psi^1_T, \\ 0 = H^1_{u^1_T} u^1_T + H^1_{x_T} x_T + H^1_{u^2_T} u^2_T + \bar{h}^1_T, \\ 0 = H^1_{x_{T+1}} x_{T+1} + \bar{h}^1_{T+1}, \\ 0 = \gamma^1_{T:T+1} \odot s^1_{T:T+1} - \rho \mathbf{1}, \\ 0 = G^1_{u^1_T} u^1_T + G^1_{x_T} x_T + G^1_{u^2_T} u^2_T + \bar{g}^1_T - s^1_T, \\ 0 = G^1_{x_{T+1}} x_{T+1} + \bar{g}^1_{T+1} - s^1_{T+1}, \\ 0 = K^2_{T,\rho}(\mathbf{z}^2_T). \end{cases}$$

We construct the KKT conditions $0 = K^2_{t,\rho}(\mathbf{z}^2_t)$ of player 2 at time $t < T$:

$$\begin{cases} 0 = \Sigma^2_{j=1} R^{2,2,j}_t u^j_t + S^{2,2}_t x_t + r^{2,2}_t + B^{2\top}_t \lambda^2_t - G^{2\top}_{u^2_t} \gamma^2_t - H^{2\top}_{u^2_t} \mu^2_t, \\ 0 = Q^2_{t+1} x_{t+1} + q^2_{t+1} - \lambda^2_t - G^{2\top}_{x_{t+1}} \gamma^2_{t+1} - H^{2\top}_{x_{t+1}} \mu^2_{t+1} \\ \quad - A^\top_{t+1} \lambda^2_{t+1} + \Sigma^2_{j=1} S^{2,j}_{t+1} u^j_{t+1} + (\nabla_{x_{t+1}} \pi^1_{t+1,\rho})^\top \eta^2_t, \\ 0 = \Sigma^2_{j=1} R^{2,1,j}_{t+1} u^j_{t+1} + S^{2,1}_{t+1} x_{t+1} + r^{2,1}_{t+1} + B^{1\top}_{t+1} \lambda^2_{t+1} - G^{2\top}_{u^1_{t+1}} \gamma^2_{t+1} - H^{2\top}_{u^1_{t+1}} \mu^2_{t+1} - \eta^2_t, \\ 0 = x_{t+1} - A_t x_t - B^1_t u^1_t - B^2_t u^2_t - c_t, \\ 0 = H^2_{u^2_t} u^2_t + H^2_{x_t} x_t + H^2_{u^1_t} u^1_t + \bar{h}^2_t \\ 0 = \gamma^2_t \odot s^2_t - \rho \mathbf{1}, \\ 0 = G^2_{u^2_t} u^2_t + G^2_{x_t} x_t + G^2_{u^1_t} u^1_t + \bar{g}^2_t - s^2_t, \\ 0 = K^1_{t+1,\rho}(\mathbf{z}^1_{t+1}). \end{cases}$$

We construct the KKT conditions $0 = K^1_{t,\rho}(\mathbf{z}^1_t)$ of player 1 at time $t < T$:

$$\begin{cases} 0 = \Sigma_{j=1}^2 R_t^{1,1,j} u_t^j + S_t^{1,1} x_t + r_t^{1,1} + B_t^{1\top} \lambda_t^1 - G_{u_t^1}^{1\top} \gamma_t^1 - H_{u_t^1}^{1\top} \mu_t^1 + (\nabla_{u_t^1} \pi_{t,\rho}^2)^\top \psi_t^1, \\ 0 = Q_{t+1}^1 x_{t+1} + q_{t+1}^1 - \lambda_t^1 - G_{x_{t+1}}^{1\top} \gamma_{t+1}^1 - H_{x_{t+1}}^{1\top} \mu_{t+1}^1 \\ \quad - A_{t+1}^\top \lambda_{t+1}^1 + \Sigma_{j=1}^2 S_{t+1}^{1,j} u_{t+1}^j + (\nabla_{x_{t+1}} \pi_{t+1,\rho}^2)^\top \eta_t^1, \\ 0 = \Sigma_{j=1}^2 R_t^{1,2,j} u_t^j + S_t^{1,2} x_t + r_t^{1,2} + B_t^{2\top} \lambda_t^1 - G_{u_t^2}^{1\top} \gamma_t^1 - H_{u_t^2}^{1\top} \mu_t^1 - \psi_t^1, \\ 0 = \Sigma_{j=1}^2 R_{t+1}^{1,2,j} u_t^j + S_{t+1}^{1,2} x_{t+1} + r_{t+1}^{1,2} + B_{t+1}^{2\top} \lambda_{t+1}^1 - G_{u_{t+1}^2}^{1\top} \gamma_{t+1}^1 - H_{u_{t+1}^2}^{1\top} \mu_{t+1}^1 - \eta_t^1, \\ 0 = H_{u_t^1}^1 u_t^1 + H_{x_t}^1 x_t + H_{u_t^2} u_t^2 + \bar{h}_t^1, \\ 0 = \gamma_t^1 \odot s_t^1 - \rho \mathbf{1}, \\ 0 = G_{u_t^1}^1 u_t^1 + G_{x_t}^1 x_t + G_{u_t^2}^1 u_t^2 + \bar{g}_t^1 - s_t^1, \\ 0 = K_{t,\rho}^2(\mathbf{z}_t^2). \end{cases}$$

We continue the above construction process until $i = 1$ and $t = 0$.

Acknowledgments. We would like to thank Professor Lillian Ratliff, Professor Forrest Laine, and Eli Brock for valuable discussions and comments. We used ChatGPT-4 [38] to check the grammar in sections 1 and 2.

REFERENCES

[1] Y. Bai, C. Jin, H. Wang, and C. Xiong, *Sample-efficient learning of Stackelberg equilibria in general-sum games*, in Advances in Neural Information Processing Systems 34, Curran Associates, 2021, pp. 25799–25811.

[2] L. Bakule and M. Straškraba, *On structural control strategies in aquatic ecosystems*, Ecol. Model., 39 (1987), pp. 171–180, https://doi.org/10.1016/0304-3800(87)90019-6.

[3] R. G. Bartle and J. T. Joichi, *The preservation of convergence of measurable functions under composition*, Proc. Amer. Math. Soc., 12 (1961), pp. 122–126, https://doi.org/10.1090/S0002-9939-1961-0120342-2.

[4] T. Basar, *On the uniqueness of the Nash solution in linear-quadratic differential games*, Internat. J. Game Theory, 5 (1976), pp. 65–90, https://doi.org/10.1007/BF01753310.

[5] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, Classics Appl. Math. 23, SIAM, Philadelphia, 1999, https://doi.org/10.1137/1.9781611971132.

[6] R. Bellman, *Dynamic Programming*, Technical report, Rand Corp., Santa Monica, CA, 1956.

[7] A. Bensoussan, S. Chen, A. Chutani, S. P. Sethi, C. C. Siu, and S. C. P. Yam, *Feedback Stackelberg–Nash equilibria in mixed leadership games with an application to cooperative advertising*, SIAM J. Control Optim., 57 (2019), pp. 3413–3444, https://doi.org/10.1137/17M1153212.

[8] T. Besselmann, J. Lofberg, and M. Morari, *Explicit MPC for LPV systems: Stability and optimality*, IEEE Trans. Automat. Control, 57 (2012), pp. 2322–2332, https://doi.org/10.1109/TAC.2012.2187400.

[9] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[10] O. Chen and M. Ben-Akiva, *Game-theoretic formulations of interaction between dynamic traffic control and dynamic traffic assignment*, Transp. Res. Record, 1617 (1998), pp. 179–188, https://doi.org/10.3141/1617-25.

[11] R. Chinchilla, G. Yang, and J. P. Hespanha, *Newton and interior-point methods for (constrained) nonconvex–nonconcave minmax optimization with stability and instability guarantees*, Math. Control Signals Systems, 36 (2024), pp. 381–421, https://doi.org/10.1007/s00498-023-00371-4.

[12] E. Dockner, *Differential Games in Economics and Management Science*, Cambridge University Press, 2000.

[13] F. Fabiani, M. A. Tajeddini, H. Kebriaei, and S. Grammatico, *Local Stackelberg equilibrium seeking in generalized aggregative games*, IEEE Trans. Automat. Control, 67 (2021), pp. 965–970, https://doi.org/10.1109/TAC.2021.3077874.

[14] P. Franceschi, N. Pedrocchi, and M. Beschi, *Human-robot role arbitration via differential game theory*, IEEE Trans. Autom. Sci. Eng., 21 (2024), pp. 5953–5968, https://doi.org/10.1109/TASE.2023.3320708.

[15] B. Gardner and J. Cruz, *Feedback Stackelberg strategy for M-level hierarchical games*, IEEE Trans. Automat. Control, 23 (1978), pp. 489–491, https://doi.org/10.1109/TAC.1978.1101733.

[16] N. Goswami, S. K. Mondal, and S. Paruya, *A comparative study of dual active-set and primal-dual interior-point method*, IFAC Proc. Vol., 45 (2012), pp. 620–625, https://doi.org/10.3182/20120710-4-SG-2026.00029.

[17] X. He, A. Prasad, S. P. Sethi, and G. J. Gutierrez, *A survey of Stackelberg differential game models in supply and marketing channels*, J. Syst. Sci. Syst. Eng., 16 (2007), pp. 385–413, https://doi.org/10.1007/s11518-007-5058-2.

[18] Y.-C. Ho, P. Luh, and R. Muralidharan, *Information structure, Stackelberg games, and incentive controllability*, IEEE Trans. Automat. Control, 26 (1981), pp. 454–460, https://doi.org/10.1109/TAC.1981.1102652.

[19] H. Hu, K. Nakamura, K.-C. Hsu, N. E. Leonard, and J. F. Fisac, *Emergent coordination through game-induced nonlinear opinion dynamics*, in 2023 62nd IEEE Conference on Decision and Control (CDC), IEEE, 2023, pp. 8122–8129.

[20] R. Janin, *Directional Derivative of the Marginal Function in Nonlinear Programming*, Springer, 1984.

[21] M. Jungers, *Feedback strategies for discrete-time linear-quadratic two-player descriptor games*, Linear Algebra Appl., 440 (2014), pp. 1–23, https://doi.org/10.1016/j.laa.2013.10.050.

[22] H. Khan and D. Fridovich-Keil, *Leadership inference for multi-agent interactions*, IEEE Robotics Autom Lett., 9 (2024), pp. 4671–4678, https://doi.org/10.1109/LRA.2024.3381469.

[23] M. Kimmel and S. Hirche, *Invariance control for safe human–robot interaction in dynamic environments*, IEEE Trans. Robot., 33 (2017), pp. 1327–1342, https://doi.org/10.1109/TRO.2017.2750697.

[24] D. Korzhyk, Z. Yin, C. Kiekintveld, V. Conitzer, and M. Tambe, *Stackelberg vs. Nash in security games: An extended investigation of interchangeability, equivalence, and uniqueness*, J. Artificial Intelligence Res., 41 (2011), pp. 297–327, https://doi.org/10.1613/jair.3269.

[25] F. Laine, D. Fridovich-Keil, C.-Y. Chiu, and C. Tomlin, *The computation of approximate generalized feedback Nash equilibria*, SIAM J. Optim., 33 (2023), pp. 294–318, https://doi.org/10.1137/21M142530X.

[26] S. M. LaValle and S. Hutchinson, *Game theory as a unifying structure for a variety of robot tasks*, in Proceedings of 8th IEEE International Symposium on Intelligent Control, IEEE, 1993, pp. 429–434.

[27] S. Le Cleac'h, M. Schwager, and Z. Manchester, *ALGAMES: A fast augmented Lagrangian solver for constrained dynamic games*, Auton. Robot., 46 (2022), pp. 201–215, https://doi.org/10.1007/s10514-021-10024-7.

[28] J. Li, C.-Y. Chiu, L. Peters, S. Sojoudi, C. Tomlin, and D. Fridovich-Keil, *Cost inference for feedback dynamic games from noisy partial state observations and incomplete trajectories*, in Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, 2023, pp. 1062–1070.

[29] M. Li, J. Qin, and L. Ding, *Two-player Stackelberg game for linear system via value iteration algorithm*, in 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), IEEE, 2019, pp. 2289–2293.

[30] T. Li and S. P. Sethi, *A review of dynamic Stackelberg game models*, Discrete Contin. Dyn. Syst. Ser. B, 22 (2017), pp. 125–159, https://doi.org/10.3934/dcdsb.2017007.

[31] R. Lucchetti, F. Mignanego, and G. Pieri, *Existence theorems of equilibrium points in Stackelberg games with constraints*, Optimization, 18 (1987), pp. 857–866, https://doi.org/10.1080/02331938708843300.

[32] M. Maljkovic, G. Nilsson, and N. Geroliminis, *On finding the leader's strategy in quadratic aggregative stackelberg pricing games*, in 2023 European Control Conference (ECC), 2023, pp. 1–6, https://doi.org/10.23919/ECC57647.2023.10178392.

[33] G. Martín-Herrán and S. J. Rubio, *On coincidence of feedback and global Stackelberg equilibria in a class of differential games*, European J. Oper. Res., 293 (2021), pp. 761–772, https://doi.org/10.1016/j.ejor.2020.12.022.

[34] S. Mondal and P. V. Reddy, *Linear quadratic Stackelberg difference games with constraints*, in 2019 18th European Control Conference (ECC), IEEE, 2019, pp. 3408–3413.

[35] T. Mylvaganam and A. Astolfi, *Approximate solutions to a class of nonlinear Stackelberg differential games*, in 53rd IEEE Conference on Decision and Control, IEEE, 2014, pp. 420–425.
[36] J. Nash, *Non-cooperative games*, Ann. of Math. (2), 54 (1951), pp. 286–295, https://doi.org/10.2307/1969529.
[37] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 1999.
[38] OpenAI, *ChatGPT*-4, software, 2023, https://openai.com/chatgpt.
[39] L. Peters, D. Fridovich-Keil, C. J. Tomlin, and Z. N. Sunberg, *Inference-based strategy alignment for general-sum differential games*, in Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, 2020, pp. 1037–1045.
[40] P. V. Reddy and G. Zaccour, *Feedback Nash equilibria in linear-quadratic difference games with constraints*, IEEE Trans. Automat. Control, 62 (2016), pp. 590–604, https://doi.org/10.1109/TAC.2016.2555879.
[41] W. Rudin, et al., *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill, New York, 1976.
[42] M. Simaan and J. B. Cruz, Jr., *Additional aspects of the Stackelberg strategy in nonzero-sum games*, J. Optim. Theory Appl., 11 (1973), pp. 613–626, https://doi.org/10.1007/BF00935561.
[43] Z. Sun, M. Greiff, A. Robertsson, and R. Johansson, *Feasible coordination of multiple homogeneous or heterogeneous mobile vehicles with various constraints*, in 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 1008–1013, https://doi.org/10.1109/ICRA.2019.8793834.
[44] A. Talebpour, H. S. Mahmassani, and S. H. Hamdar, *Modeling lane-changing behavior in a connected environment: A game theory approach*, Transp. Res. Proc., 7 (2015), pp. 420–440, https://doi.org/10.1016/j.trpro.2015.06.022.
[45] B. Tolwinski, *A Stackelberg solution of dynamic games*, IEEE Trans. Automat. Control, 28 (1983), pp. 85–93, https://doi.org/10.1109/TAC.1983.1103139.
[46] K. G. Vamvoudakis, F. L. Lewis, M. Johnson, and W. E. Dixon, *Online learning algorithm for Stackelberg games in problems with hierarchy*, in 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), IEEE, 2012, pp. 1883–1889.
[47] H. von Stackelberg, *The Theory of the Market Economy*, Oxford University Press, 1952.
[48] K. Wang, L. Xu, A. Perrault, M. K. Reiter, and M. Tambe, *Coordinating followers to reach better equilibria: End-to-end gradient descent for Stackelberg games*, in Proceedings of the 36th AAAI Conference on Artificial Intelligence, 2022, pp. 5219–5227.
[49] E. R. Weintraub, *On the existence of a competitive equilibrium:* 1930-1954, J. Econ. Lit., 21 (1983), pp. 1–39.
[50] D. Xie, *On time inconsistency: A technical issue in Stackelberg differential games*, J. Econ. Theory, 76 (1997), pp. 412–430, https://doi.org/10.1006/jeth.1997.2308.
[51] K. Xu, X. Zhao, and X. Han, *Adaptive dynamic programming for a class of two-player Stackelberg differential games*, in 2020 International Conference on System Science and Engineering (ICSSE), IEEE, 2020, pp. 1–6.
[52] J. H. Yoo and R. Langari, *A Stackelberg game theoretic driver model for merging*, in ASME 2013 Dynamic Systems and Control Conference, Vol. 2, American Society of Mechanical Engineers, 2013, V002T30A003.
[53] Y. Zhao and Q. Zhu, *Stackelberg game-theoretic trajectory guidance for multi-robot systems with Koopman operator*, in 2024 IEEE International Conference on Robotics and Automation (ICRA), 2024, pp. 12326–12332, https://doi.org/10.1109/ICRA57147.2024.10610940.