Journal of Dynamics and Games
©American Institute of Mathematical Sciences
Volume **8**, Number **4**, October **2021**
doi:10.3934/jdg.2021023
pp. **403–443**

LINEAR-QUADRATIC ZERO-SUM MEAN-FIELD TYPE GAMES: OPTIMALITY CONDITIONS AND POLICY OPTIMIZATION

René Carmona, Kenza Hamidouche, Mathieu Laurière and Zongjun Tan

Department of Operations Research and Financial Engineering
Princeton University
Princeton, NJ 08540, USA

Abstract. In this paper, zero-sum mean-field type games (ZSMFTG) with linear dynamics and quadratic cost are studied under infinite-horizon discounted utility function. ZSMFTG are a class of games in which two decision makers whose utilities sum to zero, compete to influence a large population of indistinguishable agents. In particular, the case in which the transition and utility functions depend on the state, the action of the controllers, and the mean of the state and the actions, is investigated. The optimality conditions of the game are analysed for both open-loop and closed-loop controls, and explicit expressions for the Nash equilibrium strategies are derived. Moreover, two policy optimization methods that rely on policy gradient are proposed for both model-based and sample-based frameworks. In the model-based case, the gradients are computed exactly using the model, whereas they are estimated using Monte-Carlo simulations in the sample-based case. Numerical experiments are conducted to show the convergence of the utility function as well as the two players' controls.

1. **Introduction.** Decision making in multi-agent systems has recently received an increasing interest from both theoretical and empirical viewpoints. For instance, multi-agent reinforcement learning (MARL) has been applied successfully to problems ranging from self-driving cars and robotics to games, while game-theoretic models have been exploited to study several prominent decision-making problems in engineering, economics and finance.

In multi-agent systems, a large number of interacting agents either cooperate or compete to optimize a certain individual or common goal. MARL and stochastic games were shown to model well systems with a small number of agents. However, as the number of agents becomes large, analysing such systems becomes intractable due to the exponential growth of agent interactions and the prohibitive computational cost. To tackle this issue, mean-field approximations, borrowed from statistical physics, were considered to study the limit behaviour of systems in which the agents are indistinguishable and their decisions are influenced by the empirical distribution of the other agents.

2020 *Mathematics Subject Classification.* Primary: 91A05, 91A07, 93E20, 49N80.

Key words and phrases. Mean field games, mean field control, mean field type games, zero sum games.

A preliminary version of this work was submitted to the 59th Conference on Decision and Control.

Mean-field games (MFGs) [42, 46] and their variants mean-field type control (MFC) [14] and mean-field type games (MFTG) [10] consist of studying the global behaviour of systems composed of infinitely many agents which interact in a symmetric manner. In particular, the mean-field approximation captures all agent-to-agent interactions that, individually, have a negligible influence on the overall system's evolution.

An MFG corresponds to the asymptotic limit of the situation in which all the agents compete to minimize their individual utility. In this case, the solution concept is a Nash equilibrium, in which a typical agent is worse-off if she deviates unilaterally. From the point of view of the global system, a better solution can be found by a central planner who tries to minimize the social utility by prescribing the control that each agent should use. This leads to the notion of MFC, which can be viewed as the optimal control of a McKean-Vlasov (MKV) dynamics, in which the evolution of the state process is influenced by its own distribution. Last, mean-field type games are a framework that models control problems involving several decision makers and mean-field interactions. Typical motivations are problems in which large coalitions compete or in which several agents try to influence a large population [16,34]. These three types of models have found numerous applications [12], e.g. in finance [21], energy production [8,13], crowd motion [3,6], wireless communications [44, 50, 53], distributed robotics [47] and systemic risk [24, 36].

In the past decade, many contributions have contributed to develop the theory of such mean-field problems. In order to study their solutions, a key point is the derivation of optimality conditions, which are typically phrased either in terms of partial differential equations (PDEs) or in terms of forward-backward stochastic differential equations (FBSDEs). For a detailed account, see e.g. [15,20,23] and the references therein. As a cornerstone for applications, the development of numerical methods for these mean-field problems has also attracted a growing interest. Assuming full knowledge of the model, methods for which convergence guarantees have been established include finite difference schemes for partial differential equations [1, 2], semi-Lagrangian schemes [22], augmented Lagrangian or primal-dual methods [5,17,18], value iteration algorithm [9], or neural network based stochastic methods [26, 27]; see e.g. [4] for a recent overview. However, in many practical situations, the model is not fully known and these methods can not be employed. Hence model-free or sample-based methods, in which the optimization is performed while having only access to a simulator instead of knowing the model, have recently been investigated. For mean-field games, fixed-point [41], fictitious play scheme [37] or actor-critic method [39] have been combined with model-free methods to compute the best response, whereas for mean-field control problems, the solution has been approximated using policy gradient [28] or Q-learning [29, 40]. Despite recent progress, these methods remain restricted to mean-field problems with simple structures which have a common point: the decision makers are either infinitesimal and identical players or a single central planner. More complex models are often needed to tackle applications, such as settings in which a mean-field dynamics is influenced by several distinguishable decision makers. Such situations can typically be modeled by a MFTG.

An archetypal MFTG is the case of mean-field zero-sum games. Two-player zero-sum games in their standard stochastic form, with no mean-field interactions, have been extensively studied in the literature [55]. In this class of games, two decision makers compete to respectively maximize and minimize the same utility function.

The large literature on this topic is motivated by many applications and by connections with robust control [11]. Recently, generalizations to the case where the state dynamics is of MKV type have been introduced in continuous time over a finite time horizon. Optimality conditions have been derived using the theory of backward stochastic differential equations (BSDEs) in [56], using the dynamic programming principle and partial differential equations (PDEs) in [31] or using a weak formulation in [33]. All these works assume that the controls take values in a compact space, and hence are not applicable to a general linear-quadratic setting. Along a different line, zero-sum games with mean-field interactions have also attracted interest for their connections with generative adversarial nets (GANs) [19, 35].

Although general stochastic problems with mean-field interactions can be studied from a theoretical perspective, explicit computation of the solution and numerical illustration of the Nash equilibrium are challenging. In standard optimal control, linear-quadratic (LQ) models, where the dynamics are linear and the cost is quadratic, usually have analytical or easily tractable solutions, which makes them very popular. These problems have also been considered in the optimization and machine learning communities, since algorithms with proof of convergence can be developed, see e.g. [38] where the authors prove convergence of model-based and sample-based policy gradient methods for a LQ optimal control problem. Sample-based methods have also been used to solve (standard) LQ zero-sum games. In [7], a discrete-time linear quadratic zero-sum game with infinite time horizon is studied and a Q-learning algorithm is proposed, which is proved to converge to the Nash equilibrium. In [57], the authors study LQ zero-sum games and propose three projected nested-gradient methods that are shown to converge to the Nash equilibrium of the game. However, none of these contributions tackle mean-field interactions in a zero-sum setting.

In the present work, under a discrete time, infinite-horizon and discounted utility function, we investigate zero-sum mean-field type games (ZSMFTG) of linear-quadratic type, which, to the best of our knowledge, had not been the focus of any work before. In particular, we address the case in which the transition and utility functions do not only depend on the state and the action of the controllers, but also the mean of the state and the actions. Moreover, the state is subject to a common noise. The structure of the problem and the infinite horizon regime allow us to identify the form of the equilibrium controls as linear combinations of the state and its mean conditioned on the common noise, both in the open-loop and the closed-loop settings. To learn the equilibrium, we extend the policy-gradient techniques developed in [28] for MFC, to the ZSMFTG framework. We design policy optimization methods in which the gradients are either computed exactly using the LQ model or estimated using Monte-Carlo samples when the model is not fully known.

The rest of the paper is organized as follows. In Section 2, the zero-sum mean-field type game is formulated, preceded by a N-agent control problem which motivates this setting. In Section 3, we present the rigorous probabilistic setup for the zero-sum mean-field type game under consideration. Open-loop controls are investigated in Section 4. After defining the set of admissible controls, we prove a Pontryagin maximum principle giving necessary and sufficient conditions of optimality, see Propositions 12 and 14. Section 5 considers closed-loop controls which are linear in the state and the mean. Focusing on the coefficients of the linear

combination, we define a notion of admissible controls and prove sufficient conditions of optimality, see Proposition 32 and Corollary 34. The connection between equilibria in the open-loop and the closed-loop information structures are studied in Section 6, see Lemma 36 and Remark 37. Focusing on closed-loop controls, expressions for the gradient of the utility function and a necessary condition of optimality are derived in Section 7, and both model-based and model-free policy optimization methods are proposed. In Subsection 7.4, we report numerical experiments to show the convergence of the controls and the utility function. Section 8 concludes the paper.

2. Model and problem formulation.

In this section, we first present a zero-sum game in which two controllers compete to influence a population of agents. The agents interact in a symmetric way, through the empirical distribution of their states and actions. We then present a mean-field version of the game (corresponding to the situation where $N \to +\infty$), in which the two controllers influence a state whose dynamics is of MKV type.

2.1. N-agent problem.

Consider a system composed of a population $\{1, \dots, N\}$ with N indistinguishable *agents*. We investigate the case in which these agents have symmetric interactions and are influenced by two *decision makers*, also called *controllers* or *players*, competing to optimize a criterion. In particular, we are interested in the linear-quadratic zero-sum case. Here, the state evolution of an agent $i \in \{1, \dots, N\}$ is given by

$$x^i_{t+1} = Ax^i_t + \bar{A}\bar{x}_t + B_1 u^i_{1,t} + \bar{B}_1 \bar{u}_{1,t} + B_2 u^i_{2,t} + \bar{B}_2 \bar{u}_{2,t} + \epsilon^i_{t+1} + \epsilon^0_{t+1}, \quad (1)$$

with initial condition $x^i_0 = \epsilon^i_0 + \epsilon^0_0$, where x^i_0 is the initial state of agent i to which we introduce randomness with ϵ^i_0 and ϵ^0_0. At each time t, $x^i_t \in \mathbb{R}^d$ corresponds to the state of the i-th agent in the population, and $u^i_{1,t} \in \mathbb{R}^\ell$ and $u^i_{2,t} \in \mathbb{R}^\ell$ are the controls prescribed to this agent respectively by the first and the second decision maker. The noise terms ϵ^0_{t+1} and ϵ^i_{t+1} are independent of each other and of ϵ^0_0 and ϵ^i_0, and we assume they have a finite second moment. Moreover, the noise terms ϵ^0_{t+1} for $t \geq 0$ are assumed to be identically distributed with mean 0, and similarly for ϵ^i_{t+1} for $t \geq 0$. The interpretation of the noise terms is that ϵ^0_t is a common noise affecting the position of all the agents, whereas ϵ^i_t is an idiosyncratic noise affecting only the position of the i-th agent. $A, \bar{A}, B_i, \bar{B}_i$ are fixed matrices with suitable dimensions. Here, $\bar{x}_t = \frac{1}{N}\sum_{i=1}^N x^i_t$, is the sample average of the individual states, and similarly for u_1 and u_2: $\bar{u}_{j,t} = \frac{1}{N}\sum_{i=1}^N u^i_{j,t}$. The instantaneous utility is defined by

$$\begin{aligned} c(x, \bar{x}, \mathrm{u}_1, \bar{\mathrm{u}}_1, \mathrm{u}_2, \bar{\mathrm{u}}_2) = {} & (x - \bar{x})^\top Q (x - \bar{x}) + \bar{x}^\top (Q + \bar{Q}) \bar{x} \\ & + (\mathrm{u}_1 - \bar{\mathrm{u}}_1)^\top R_1 (\mathrm{u}_1 - \bar{\mathrm{u}}_1) + \bar{\mathrm{u}}_1^\top + (R_1 + \bar{R}_1) \bar{\mathrm{u}}_1 \\ & - (\mathrm{u}_2 - \bar{\mathrm{u}}_2)^\top R_2 (\mathrm{u}_2 - \bar{\mathrm{u}}_2) - \bar{\mathrm{u}}_2^\top (R_2 + \bar{R}_2) \bar{\mathrm{u}}_2, \end{aligned} \quad (2)$$

where $Q, \bar{Q}, R_i, \bar{R}_i$ are deterministic symmetric matrices of suitable sizes such that $R_i, R_i + \bar{R}_i$ for $i = 1, 2$ are positive definite.

The objective of each controller in this zero-sum problem is to minimize (resp. maximize) the N-agent utility functional

$$J^N(\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2) = \mathbb{E}\left[\sum_{t=0}^{+\infty} \gamma^t \bar{c}^N(\underline{x}_t, \underline{u}_{1,t}, \underline{u}_{2,t})\right],$$

where $\underline{x}_t = (x_t^1, \dots, x_t^N)$, and $\underline{\mathbf{u}}_i = (\underline{u}_{i,t})_t$ with $\underline{u}_{i,t} = (u_{i,t}^1, \dots, u_{i,t}^N)$ (we use a boldface to denote a function of time and an underline to denote a vector of size N), and $\bar{c}^N$ is the average utility, defined by

$$\bar{c}^N(\underline{x}_t, \underline{u}_{1,t}, \underline{u}_{2,t}) = \frac{1}{N}\sum_{i=1}^{N} c(x_t^i, \bar{x}_t, u_{1,t}^i, \bar{u}_{1,t}, u_{2,t}^i, \bar{u}_{2,t}).$$

Remark 1. An interesting special case is the situation in which each decision maker controls a different population. This corresponds to a zero-sum game between two large coalitions. This setting can be covered in the following way. Assume that $d = 2d'$ for some integer d'. Consider, for the dynamics, block matrices of the form:

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad B_1 = \begin{pmatrix} B_1^1 \\ 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 \\ B_2^2 \end{pmatrix},$$

and

$$\bar{A} = \begin{pmatrix} \bar{A}_{11} & \bar{A}_{12} \\ \bar{A}_{21} & \bar{A}_{22} \end{pmatrix}, \quad \bar{B}_1 = \begin{pmatrix} \bar{B}_1^1 \\ \bar{B}_1^2 \end{pmatrix}, \quad \bar{B}_2 = \begin{pmatrix} \bar{B}_2^1 \\ \bar{B}_2^2 \end{pmatrix}.$$

Then the dynamics (1) rewrites, with the notation $x = (x_1, x_2)$ where $x_i \in \mathbb{R}^{d'}$ and similarly for ϵ^0, ϵ^i,

$$\begin{aligned} dx_{1,t}^i &= \left[A_1 x_{1,t}^i + \bar{A}_{11}\bar{x}_{1,t} + \bar{A}_{12}\bar{x}_{2,t} + B_1^1 u_{1,t}^i + \bar{B}_1^1\bar{u}_{1,t} + \bar{B}_1^2\bar{u}_{2,t}\right] + \epsilon_{1,t}^i + \epsilon_{1,t}^0, \\ dx_{2,t}^i &= \left[A_2 x_{2,t}^i + \bar{A}_{21}\bar{x}_{1,t} + \bar{A}_{22}\bar{x}_{2,t} + B_2^2 u_{2,t}^i + \bar{B}_1^2\bar{u}_{1,t} + \bar{B}_2^2\bar{u}_{2,t}\right] + \epsilon_{2,t}^i + \epsilon_{2,t}^0. \end{aligned}$$

Note that the evolution of the two halves of vector x are coupled only through their expectations and the expectation of the control used for the other half. We can thus interpret each half as the state of a player in a different population where each population has N indistinguishable agents.

2.2. **Mean-field problem.** Here, we consider the limit of the N-agent case. The dynamics is given by

$$x_{t+1} = Ax_t + \bar{A}\bar{x}_t + B_1 u_{1,t} + \bar{B}_1\bar{u}_{1,t} + B_2 u_{2,t} + \bar{B}_2\bar{u}_{2,t} + \epsilon_{t+1}^0 + \epsilon_{t+1}^1, \quad (3)$$

with initial condition

$$x_0 = \epsilon_0^0 + \epsilon_0^1.$$

Here and thereafter, when considering the mean-field problem, we use the notation $\bar{x}_t = \mathbb{E}[x_t|(\epsilon_s^0)_{0\le s\le t}]$ for the expectation of the state conditional on the realization of the common noise, and likewise for $\mathbf{u}_1$ and $\mathbf{u}_2$. Note that (3) is a dynamics of MKV type since it is influenced by its own distribution and by the distribution of the actions. The utility function takes the form

$$J(\mathbf{u}_1, \mathbf{u}_2) = \mathbb{E}\Big[\sum_{t=0}^{+\infty} \gamma^t c_t\Big], \quad (4)$$

where $\gamma \in [0, 1]$ is a discount factor, and the instantaneous utility at time t is defined as

$$c_t = c(x_t, \bar{x}_t, u_{1,t}, \bar{u}_{1,t}, u_{2,t}, \bar{u}_{2,t}), \quad (5)$$

where the function c is as in the N-agent problem.

The goal is to find an equilibrium in the sense of Nash, namely a situation in which none of the controllers can benefit from a unilateral deviation. Such problems are usually framed as min max (or max min) games. Due to subtle questions of

control admissibility, we will view an equilibrium as a saddle point (see Definitions 7 and 21 below for the open-loop and closed-loop settings respectively).

Our framework is a generalization of the mean-field control setup, in which there is a single decision maker. It can also be viewed as a variant of a Nash mean-field control setup studied in [16] or a mean-field type game [34] in which several mean-field decision makers compete in a general-sum game.

Next, we study the existence of the Nash equilibrium and derive its closed-form expression for the formulated ZSMFTG.

3. Probabilistic setup.

In this section we rigorously define the model of MKV dynamics with common noise. It is analogous to the one considered in [28], except for the fact that there are two decision makers instead of one. A convenient way to think about this model is to view the state x_t of the system at time t as a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = \Omega^0 \times \Omega^1$, $\mathcal{F} = \mathcal{F}^0 \times \mathcal{F}^1$ and $\mathbb{P} = \mathbb{P}^0 \times \mathbb{P}^1$. In this set-up, if $\omega = (\omega^0, \omega^1)$, $\epsilon^0_t(\omega) = \tilde{\epsilon}^0_t(\omega^0)$ and $\epsilon^1_t(\omega) = \tilde{\epsilon}^1_t(\omega^1)$ where $(\tilde{\epsilon}^0_t)_{t=1,2,\dots}$ and $(\tilde{\epsilon}^1_t)_{t=1,2,\dots}$ are i.i.d. sequences of mean-zero random variables on $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ and $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ respectively, while the initial sources of randomness $\tilde{\epsilon}^0_0$ and $\tilde{\epsilon}^1_0$ are random variables on $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ and $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ with distributions μ^0_0 and μ^1_0 respectively, which are independent of each other and independent of $(\tilde{\epsilon}^0_t)_{t=1,2,\dots}$ and $(\tilde{\epsilon}^1_t)_{t=1,2,\dots}$. We denote by $\mathcal{F}_t$ the filtration generated by the noise up until time t, that is $\mathcal{F}_t = \sigma(\epsilon^0_0, \epsilon^1_0, \epsilon^0_1, \epsilon^1_1, \dots, \epsilon^0_t, \epsilon^1_t)$. We assume that the variance of random variables ϵ^0_t and ϵ^1_t are constant along time, and these variances are denoted by $\Sigma^0 = \mathbb{E}[(\epsilon^0_t)^\top \epsilon^0_t]$ and $\Sigma^1 = \mathbb{E}[(\epsilon^1_t)^\top \epsilon^1_t]$ for every $t \geq 1$.

At each time $t \geq 0$, x_t and $u_{i,t}$ with $i = 1, 2$ are random elements defined on $(\Omega, \mathcal{F}, \mathbb{P})$ representing the state of the system and the controls exerted by a pair of generic agents. Using the fact that the idiosyncratic noise and the common noise are independent, the quantities $\bar{x}_t$ and $\bar{u}_{i,t}$ with $i = 1, 2$ appearing in (3) are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ defined by: for $\omega = (\omega^0, \omega^1)$,

$$\bar{x}_t(\omega^0, \omega^1) = \int_{\Omega^1} x_t(\omega^0, \tilde{\omega}^1)\mathbb{P}^1(d\tilde{\omega}^1),\ \bar{u}_{i,t}(\omega^0, \omega^1) = \int_{\Omega^1} u_{i,t}(\omega^0, \tilde{\omega}^1)\mathbb{P}^1(d\tilde{\omega}^1),\ i = 1, 2.$$

Notice that $\bar{x}_t$, $\bar{u}_{1,t}$ and $\bar{u}_{2,t}$ depend only upon ω^0. In fact, the best way to think of $\bar{x}_t$ and $\bar{u}_{i,t}$ with $i = 1, 2$ is to keep in mind the following fact:

$$\bar{x}_t = \mathbb{E}[x_t|\mathcal{F}^0], \qquad \text{and} \qquad \bar{u}_{i,t} = \mathbb{E}[u_{i,t}|\mathcal{F}^0].$$

These are the mean field terms appearing in the (stochastic) dynamics of the state (3):

$$x_{t+1} = Ax_t + \bar{A}\bar{x}_t + B_1 u_{1,t} + \bar{B}_1 \bar{u}_{1,t} + B_2 u_{2,t} + \bar{B}_2 \bar{u}_{2,t} + \epsilon^0_{t+1} + \epsilon^1_{t+1}.$$

4. Open-loop information structure.

In this section, we consider open-loop controls, that is, controls available if the controllers can directly see the noise terms. We start with this class of controls because it is somehow "larger" than the class of closed-loop controls that will be considered in the next section (any closed-loop control gives rise to an open-loop control, but the converse is not always true). The main point of this section is to show that, under suitable conditions, the saddle point controls in the open-loop setting can in fact be written as linear combinations of the state and the conditional mean.

4.1. **Admissible controls.** We will use the following notation: for $n \in \mathbb{N}_+$, for any process $\mathbf{x} : \Omega \mapsto \mathbb{R}^n$,

$$\|\mathbf{x}\|_{2,\gamma} = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \|x_t\|^2\right],$$

We introduce the following sets: for $T \geq 0$, letting $\mathbb{N}_{\leq T} = \{0, \dots, T\}$,

$$\mathcal{U}_T := \Big\{\mathbf{u} : \mathbb{N}_{\leq T} \times \Omega \mapsto \mathbb{R}^\ell \mid u_t \text{ is } \mathcal{F}_t - \text{measurable}, \ \mathbb{E}\left[\sup_{t=0,\dots,T} \gamma^t \|u_t\|^2\right] < \infty\Big\},$$

$$\mathcal{U}_{loc} := \bigcup_{T \geq 0} \mathcal{U}_T, \qquad \mathcal{U} := \left\{\mathbf{u} : \mathbb{N} \times \Omega \mapsto \mathbb{R}^\ell \mid \mathbf{u} \in \mathcal{U}_{loc}, \ \|\mathbf{u}\|_{2,\gamma} < \infty\right\},$$

where we use the notation $u_t(\cdot) = \mathbf{u}(t, \cdot)$ for every $t \in \mathbb{N}$ and we identify $\mathbf{u}$ to an $\mathcal{F}$-adapted process $(u_t)_{t \geq 0}$. A process $\mathbf{u}$ is called L^2-discounted globally integrable, or L^2-integrable for short, if $\mathbf{u} \in \mathcal{U}$. Also, for $T \geq 0$,

$$\mathcal{X}_T := \Big\{\mathbf{x} : \mathbb{N}_{\leq T} \times \Omega \mapsto \mathbb{R}^d \mid x_t \text{ is } \mathcal{F}_t - \text{measurable}, \ \mathbb{E}\Big[\sup_{t=0,\dots,T} \gamma^t \|x_t\|^2\Big] < \infty\Big\},$$

$$\mathcal{X}_{loc} := \bigcup_{T \geq 0} \mathcal{X}_T, \qquad \mathcal{X} := \left\{\mathbf{x} : \mathbb{N} \times \Omega \mapsto \mathbb{R}^d \mid \mathbf{x} \in \mathcal{X}_{loc}, \ \|\mathbf{x}\|_{2,\gamma} < \infty\right\}.$$

Similarly, we identify $\mathbf{x} \in \mathcal{X}_{loc}$ or $\mathbf{x} \in \mathcal{X}$ to an $\mathcal{F}-$adapted process $(x_t)_{t \geq 0}$ in $\mathbb{R}^d$. We also say that a state process is L^2-discounted globally integrable, or simply L^2-integrable, if $\mathbf{x} \in \mathcal{X}$. Let $\mathcal{S}^d$ stand for the set of symmetric matrices in $\mathbb{R}^{d \times d}$.

In the open-loop information structure, we consider the following subset of $\mathcal{U} \times \mathcal{U}$:

$$\mathcal{U}_{ad}^{open} := \{(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U} \times \mathcal{U} \mid (x_t^{\mathbf{u}_1, \mathbf{u}_2})_{t \geq 0} \in \mathcal{X}\}$$

where the state process $(x_t^{\mathbf{u}_1, \mathbf{u}_2})_{x \geq 0}$ follows the dynamics (3). We call every element $(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}_{ad}^{open}$ an admissible (open-loop) control pair for the two players.

The following proposition is about the L^2-integrability of the state processes.

Proposition 2. *Let us assume that the process $\boldsymbol{X} = (X_t)_{t \geq 0}$ satisfies*

$$X_{t+1} = AX_t + q_t, \qquad X_0 \sim \mu_0 \tag{6}$$

where $A \in \mathbb{R}^{d \times d}$ is a fixed matrix, $\mu_0 \in \mathcal{P}^2(\mathbb{R}^d)$ so that $\mathbb{E}[\|X_0\|^2] < \infty$, and the process $\boldsymbol{q} = (q_t)_{t \geq 0}$ satisfies: $\|\mathbf{q}\|_{2,\gamma} < +\infty$ If the matrix A satisfies $\gamma \|A\|^2 < 1$, then the process $\boldsymbol{X} = (X_t)_{t \geq 0}$ satisfies $\|\boldsymbol{X}\|_{2,\gamma} < +\infty$.

Proof. Given the assumption $\gamma \|A\|^2 < 1$, we can choose $\gamma_1 \in (\gamma, 1)$ such that $\xi := \gamma_1 \|A\|^2 < 1$. Let $\eta = \gamma / \gamma_1 < 1$. From the dynamics of state process $\boldsymbol{X}$, we have for every $t \geq 1$,

$$X_t = A^t X_0 + \sum_{j=0}^{t-1} A^{t-1-j} q_j.$$

Hence, letting $C_{A,\xi,\gamma} = \frac{\gamma^{1/2} A}{\xi^{1/2}}$,

$$\begin{aligned}
\mathbb{E}\left[\gamma^t \|X_t\|^2\right] &\leq \mathbb{E}\Big[\Big(\big\|C_{A,\xi,\gamma}^t \xi^{t/2} X_0\big\| + \sum_{j=0}^{t-1} \big\|C_{A,\xi,\gamma}^{t-1-j} \xi^{(t-1-j)/2} \gamma^{(j+1)/2} q_j\big\|\Big)^2\Big] \\
&\leq 2\eta^t \xi^t \mathbb{E}[\|X_0\|^2] + 2\mathbb{E}\Big[\Big(\sum_{j=0}^{t-1} \eta^{(t-1-j)/2} \xi^{(t-1-j)/2} \gamma^{(j+1)/2} \|q_j\|\Big)^2\Big]
\end{aligned}$$

$$\leq 2\eta^t\xi^t\mathbb{E}[\|X_0\|^2] + \frac{2}{1-\eta}\Big(\sum_{j=0}^{t-1}\xi^{t-1-j}\gamma^{j+1}\mathbb{E}[\|q_j\|^2]\Big).$$

Finally, summing over t and interchanging the two summations we get:

$$\sum_{t=0}^{\infty}\mathbb{E}\left[\gamma^t\|X_t\|^2\right] \leq \frac{2}{1-\xi\eta}\mathbb{E}[\|X_0\|^2] + \frac{2\gamma}{(1-\eta)(1-\xi)}\sum_{j=0}^{\infty}\mathbb{E}\left[\gamma^j\|q_j\|^2\right] < \infty.$$

□

Note the following link with L^2-asymptotical stability: If $\mathbf{x} \in \mathcal{X}$, then we have $\lim_{t\to\infty}\mathbb{E}[\gamma^t\|x_t\|^2] = 0$.

We have the following two lemmas related to the L^2-discounted globally integrability for processes $(x_t - \bar{x}_t)_{t\geq 0}$ and $(\bar{x}_t)_{t\geq 0}$.

Lemma 3. *A process $\mathbf{x} \in \mathcal{X}$ if and only if both processes $(x_t - \bar{x}_t)_{t\geq 0} \in \mathcal{X}$ and $(\bar{x}_t)_{t\geq 0} \in \mathcal{X}$. Similarly, a control process $\mathbf{u} \in \mathcal{U}$ if and only if both processes $(u_t - \bar{u}_t)_{t\geq 0} \in \mathcal{U}$ and $(\bar{u}_t)_{t\geq 0} \in \mathcal{U}$.*

Using Proposition 2 and Lemma 3, we deduce the following result.

Lemma 4. *For any given $(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}\times\mathcal{U}$, let $(q_t^{(y)})_{t\geq 0}$ and $(q_t^{(z)})_{t\geq 0}$ be two processes in $\mathcal{U}$ given by: for every $t \geq 0$,*

$$\begin{cases} q_t^{(y)} = B_1(u_{1,t} - \bar{u}_{1,t}) + B_2(u_{2,t} - \bar{u}_{2,t}), \\ q_t^{(z)} = (B_1 + \bar{B}_1)\bar{u}_{1,t} + (B_2 + \bar{B}_2)\bar{u}_{2,t}. \end{cases} \tag{7}$$

We have:

- *If $\gamma\|A\|^2 < 1$, then the state process $\mathbf{y} = (y_t)_{t\geq 0}$ following the dynamics*

$$y_{t+1} = Ay_t + q_t^{(y)} + \epsilon_{t+1}^1, \qquad y_0 \sim \mu_0^1 \tag{8}$$

  is L^2-discounted globally integrable;
- *If $\gamma\|A + \bar{A}\|^2 < 1$, the state process $\mathbf{z} = (z_t)_{t\geq 0}$ following the dynamics*

$$z_{t+1} = (A + \bar{A})z_t + q_t^{(z)} + \epsilon_{t+1}^0, \qquad z_0 \sim \mu_0^0 \tag{9}$$

  is L^2-discounted globally integrable.

Now, we are ready to provide the L^2-discounted global integrability for the state process $\mathbf{x}^{\mathbf{u}_1,\mathbf{u}_2} = (x_t^{\mathbf{u}_1,\mathbf{u}_2})_{t\geq 0}$ following dynamics (3) controlled by two processes $\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{U}$.

Assumption 1. *$\gamma\|A\|^2 < 1$ and $\gamma\|A + \bar{A}\|^2 < 1$.*

Proposition 5. *Under Assumption 1, we have $\mathcal{U}_{ad}^{open} = \mathcal{U}\times\mathcal{U}$. In particular, the set of admissible controls $\mathcal{U}_{ad}^{open}$ is convex.*

Proof. By definition, $\mathcal{U}_{ad}^{open} \subseteq \mathcal{U}\times\mathcal{U}$. For the other inclusion, let us consider a pair of control processes $(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}\times\mathcal{U}$. We know that the corresponding state process $\mathbf{x}^{\mathbf{u}_1,\mathbf{u}_2} \in \mathcal{X}_{loc}$. Taking the conditional expectation with respect to $\mathcal{F}^0$ and denoting $\bar{x}_t^{\mathbf{u}_1,\mathbf{u}_2} = \mathbb{E}[x_t^{\mathbf{u}_1,\mathbf{u}_2}|\mathcal{F}^0]$, we notice that, for every $t \geq 0$,

$$\begin{cases} x_t^{\mathbf{u}_1,\mathbf{u}_2} - \bar{x}_t^{\mathbf{u}_1,\mathbf{u}_2} = A(x_t^{\mathbf{u}_1,\mathbf{u}_2} - \bar{x}_t^{\mathbf{u}_1,\mathbf{u}_2}) + q_t^{(y)} + \epsilon_{t+1}^1 \\ \bar{x}_t^{\mathbf{u}_1,\mathbf{u}_2} = (A + \bar{A})\bar{x}_t^{\mathbf{u}_1,\mathbf{u}_2} + q_t^{(z)} + \epsilon_{t+1}^0 \end{cases}$$

where $q_t^{(y)}$ and $q_t^{(z)}$ are given by (7). Let us denote $y_t = x_t^{\mathbf{u}_1,\mathbf{u}_2} - \bar{x}_t^{\mathbf{u}_1,\mathbf{u}_2}$ and $z_t = \bar{x}_t^{\mathbf{u}_1,\mathbf{u}_2}$ for every $t \geq 0$. Under Assumption 1, by Lemma 4, we obtain that the processes $(y_t)_{t\geq 0} \in \mathcal{X}$ and $(z_t)_{t\geq 0} \in \mathcal{X}$. which implies $\mathbf{x}^{\mathbf{u}_1,\mathbf{u}_2} \in \mathcal{X}$. Thus, $\mathcal{U}\times\mathcal{U} \subseteq \mathcal{U}_{ad}^{open}$. The convexity of $\mathcal{U}_{ad}^{open}$ is a consequence of the convexity of $\mathcal{U}$. □

Remark 6. In the closed-loop information structure (c.f. Section 5), we will see that the set of closed-loop admissible policy is not convex.

Definition 7. A pair of admissible control processes $(\mathbf{u}_1^*, \mathbf{u}_2^*) \in \mathcal{U}_{ad}^{open}$ is an open-loop saddle point (OLSP for short) for the zero-sum game if for any process $\mathbf{u}_1' \in \mathcal{U}$ and $\mathbf{u}_2' \in \mathcal{U}$, we have

$$J(\mathbf{u}_1^*, \mathbf{u}_2') \leq J(\mathbf{u}_1^*, \mathbf{u}_2^*) \leq J(\mathbf{u}_1', \mathbf{u}_2^*), \tag{10}$$

where $(\mathbf{u}_1, \mathbf{u}_2) \mapsto J(\mathbf{u}_1, \mathbf{u}_2)$ is the utility function defined in equation (4).

4.2. **Equilibrium condition.** For the sake of convenience, we use the notation $\check{A} = A - I_d$ where I_d denotes the $d\times d$ identity matrix, and $\zeta = (x, \bar{x}, u_1, \bar{u}_1, u_2, \bar{u}_2)$, so that, if we define the function b by:

$$b(\zeta) = b(x, \bar{x}, u_1, \bar{u}_1, u_2, \bar{u}_2) = \check{A}x + \bar{A}\bar{x} + B_1u_1 + \bar{B}_1\bar{u}_1 + B_2u_2 + \bar{B}_2\bar{u}_2. \tag{11}$$

The state equation (3) can be rewritten as:

$$x_{t+1} - x_t = b(x_t, \bar{x}_t, u_{1,t}, \bar{u}_{1,t}, u_{2,t}, \bar{u}_{2,t}) + \epsilon_{t+1}^0 + \epsilon_{t+1}^1 = b(\zeta_t) + \epsilon_{t+1}^0 + \epsilon_{t+1}^1. \tag{12}$$

We define the Hamiltonian function h by:

$$\begin{aligned} &h(\zeta, p) \\ &= h(x, \bar{x}, u_1, \bar{u}_1, u_2, \bar{u}_2, \eta) \\ &= [\check{A}x + \bar{A}\bar{x} + B_1u_1 + \bar{B}_1\bar{u}_1 + B_2u_2 + \bar{B}_2\bar{u}_2] \cdot p + c(x, \bar{x}, u_1, \bar{u}_1, u_2, \bar{u}_2) - \delta x \cdot p \\ &= b(\zeta) \cdot p + c(\zeta) - \delta x \cdot p \end{aligned} \tag{13}$$

for $p \in \mathbb{R}^d$, where $\delta = (1-\gamma)/\gamma$ is a positive constant representing the discount rate, $\gamma \in [0,1]$ being the discount factor. Throughout, we use the notation $\cdot$ for the scalar product in Euclidean space. We will use the following property of the Hamiltonian, under the following assumption, where $\succeq 0$ (resp. $\succ$) means that the matrix is non-negative semi definite (resp. positive definite).

Lemma 8. *If $R_1 \succeq 0$, $R_1 + \bar{R}_1 \succeq 0$ (resp. $R_2 \succeq 0$, and $R_2 + \bar{R}_2 \succeq 0$), the function h is convex w.r.t. $(u_1, \bar{u}_1)$ (resp. concave w.r.t. $(u_2, \bar{u}_2)$). It is strictly convex (resp. strictly concave) if $R_1 \succ 0$ and $(R_1 + \bar{R}_1) \succ 0$ (resp. $R_2 \succ 0$ and $(R_2 + \bar{R}_2) \succ 0$).*

Proof. For the purpose of computing gradients, Hessians and partial derivatives, we treat ζ as a $(2d + 4\ell) \times 1$ column vector by specifying its definition as $\zeta = [x^\top, \bar{x}^\top, u_1^\top, \bar{u}_1^\top, u_2^\top, \bar{u}_2^\top]^\top$. Now, for every fixed $p \in \mathbb{R}^d$, we have:

$$\nabla_\zeta h(\zeta, p) = \begin{pmatrix} \partial_x h(\zeta,p) \\ \partial_{\bar{x}} h(\zeta,p) \\ \partial_{u_1} h(\zeta,p) \\ \partial_{\bar{u}_1} h(\zeta,p) \\ \partial_{u_2} h(\zeta,p) \\ \partial_{\bar{u}_2} h(\zeta,p) \end{pmatrix} = \begin{pmatrix} p^\top(\check{A} - \delta I_d) + 2(x-\bar{x})^\top Q \\ p^\top \bar{A} + 2(\bar{x}-x)^\top Q + 2\bar{x}^\top(Q+\bar{Q}) \\ p^\top B_1 + 2(u_1 - \bar{u}_1)^\top R_1 \\ p^\top \bar{B}_1 + 2(\bar{u}_1 - u_1)^\top R_1 + 2\bar{u}_1^\top(R_1 + \bar{R}_1) \\ p^\top B_2 - 2(u_2 - \bar{u}_2)^\top R_2 \\ p^\top \bar{B}_2 - 2(\bar{u}_2 - u_2)^\top R_2 - 2\bar{u}_2^\top(R_2 + \bar{R}_2) \end{pmatrix}. \tag{14}$$

It can be seen that

$$\nabla^2_{(u_1,\bar{u}_1),(u_1,\bar{u}_1)} h(\zeta, p) = \begin{pmatrix} 2R_1 & -2R_1 \\ -2R_1 & 2(2R_1 + \bar{R}_1) \end{pmatrix},$$

is non-negative definite if the inequalities $R_1 \succeq 0$ and $R_1 + \bar{R}_1 \succeq 0$ are satisfied, and positive definite if $R_1 \succ 0$ and $(R_1 + \bar{R}_1) \succ 0$. Likewise for the second order derivatives w.r.t. $(u_2, \bar{u}_2)$. □

In order to use the stochastic version of the Pontryagin maximum principle, we introduce the notion of adjoint process associated to a given admissible pair of control processes.

Definition 9. If $\mathbf{u}_i = (u_{i,t})_{t=0,1,\dots}, i = 1, 2$ is a pair of admissible control processes and $\mathbf{x} = (x_t)_{t=0,1,\dots}$ is the corresponding state process controlled by $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$, we say that an $\mathbb{R}^d$-valued $(\mathcal{F}_t)_{t\geq 0}$-adapted process $\boldsymbol{p} = (p_t)_{t=0,1,\dots}$ is an adjoint process corresponding to $\mathbf{x}$ if it satisfies:

$$p_t = \mathbb{E}\Big[p_{t+1} + \gamma\big[(\check{A}^\top - \delta I_d)p_{t+1} + 2Qx_{t+1} + \bar{A}^\top \bar{p}_{t+1} + 2\bar{Q}\bar{x}_{t+1}\big] | \mathcal{F}_t\Big], \qquad t \geq 0, \tag{15}$$

and the transversality condition:

$$\|\mathbf{p}\|_{2,\gamma} < \infty. \tag{16}$$

It will be useful to note that the above expression (15) can equivalently be written as:

$$p_t = \gamma \mathbb{E}[A^\top p_{t+1} + 2Qx_{t+1} + \bar{A}^\top \bar{p}_{t+1} + 2\bar{Q}\bar{x}_{t+1} \mid \mathcal{F}_t]. \tag{17}$$

The following result shows that combined with the admissibility of a couple of controls, our Assumption 1 automatically implies the transversality condition.

Proposition 10. *Assume that Assumption 1 holds. For every admissible pair of control processes, there exists a unique adjoint process.*

Proof. Let $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$ be an admissible pair of control processes and let $\mathbf{x} = (x_t)_{t=0,1,\dots}$ be the corresponding state process.

Uniqueness. Let $(p_t)_{t\geq 0}$ and $(p'_t)_{t\geq 0}$ be two adjoint processes corresponding to $(\mathbf{u}_1, \mathbf{u}_2)$. We first look at the corresponding conditional processes, namely, $\bar{p}_t = \mathbb{E}[p_t|\mathcal{F}^0]$, $\bar{p}'_t = \mathbb{E}[p'_t|\mathcal{F}^0]$, $t \geq 0$. Taking the conditional expectation on both sides of equation (17) for p_t and p'_t and then the difference between equations for $\bar{p}_t$ and $\bar{p}'_t$, by condition $\gamma^{1/2}\|A + \bar{A}\| < 1$ we obtain $\mathbb{E}[\|\bar{p}_t - \bar{p}'_t\|] \leq \gamma^{1/2}\mathbb{E}\big[\|\bar{p}_{t+1} - \bar{p}'_{t+1}\|\big]$. By induction, we obtain for every $0 \leq t < s$, $\gamma^{t/2}\mathbb{E}[\|\bar{p}_t - \bar{p}'_t\|] \leq \gamma^{s/2}\mathbb{E}[\|\bar{p}_s - \bar{p}'_s\|]$. By the transversality condition (16) for $(p_s)_{s\geq 0}$ and $(p'_s)_{s\geq 0}$ and the Jensen's inequality for conditional expectation, we get

$$\lim_{s\to\infty} \gamma^{s/2}\mathbb{E}\big[\|\bar{p}_s - \bar{p}'_s\|\big] \leq \lim_{s\to\infty} \Big[(\mathbb{E}[\gamma^s \|p_s\|^2])^{1/2} + (\mathbb{E}[\gamma^s \|p'_s\|^2])^{1/2}\Big] = 0.$$

Hence $\bar{p}_t = \bar{p}'_t$, $\mathbb{P}$-a.s., for all $t \geq 0$. Similarly, from (17), we deduce that $(p_t - \bar{p}_t) = (p'_t - \bar{p}'_t)$, $\mathbb{P}$-a.s., for all $t \geq 0$. Therefore $p_t = p'_t$, $\mathbb{P}$-a.s., for all $t \geq 0$.

Existence. We proceed by constructing an approximation over a finite time horizon and then passing to the limit. For every $T > 0$, define the process $(p^T_t)_{t\geq 0}$ by: for $t \geq T$, $p^T_t = 0$, and for $t = T-1, T-2, 1, \dots, 0$,

$$p^T_t = \mathbb{E}\Big[p^T_{t+1} + \gamma\big[(\check{A}^\top - \delta I_d)p^T_{t+1} + 2Qx_{t+1} + \bar{A}^\top \bar{p}^T_{t+1} + 2\bar{Q}\bar{x}_{t+1}\big] | \mathcal{F}_t\Big],$$

where $\bar{p}_t^T = \mathbb{E}[p_t^T|\mathcal{F}^0]$. By equation (17), we have: for all $t = 0, \dots, T-1$,

$$\bar{p}_t^T = \gamma\mathbb{E}[(A+\bar{A})^\top \bar{p}_{t+1}^T + 2(Q+\bar{Q})\bar{x}_{t+1}|\mathcal{F}_t], \tag{18}$$

$$p_t^T - \bar{p}_t^T = \gamma\mathbb{E}[A^\top(p_{t+1}^T - \bar{p}_{t+1}^T) + 2Q(x_{t+1} - \bar{x}_{t+1})|\mathcal{F}_t]. \tag{19}$$

We split the proof of existence into four steps. We first study the processes $(\bar{p}_t^T)_{t\geq 0}$, $T > 0$.

Claim 1. *For every $s \geq 0$, $(\bar{p}_s^T)_{T\geq s}$ is a Cauchy sequence for convergence in L^1 under the norm $\|\cdot\|$.*

This is a direct consequence of Assumption 1 and the following property, that we prove below: For every s, T_1, T_2, T_3, such that $0 \leq s \leq T_1 < \min\{T_2, T_3\}$,

$$\mathbb{E}\left[\|\bar{p}_s^{T_2} - \bar{p}_s^{T_3}\|\right] \leq 2\tilde{\gamma}^{1-s}\tilde{\eta}^{T_1-s}M, \tag{20}$$

where $\tilde{\gamma} = \gamma^{\frac{1}{2}}$, $\tilde{\eta} = \gamma^{\frac{1}{2}}\|A+\bar{A}\|$ and $M = \frac{2}{1-\tilde{\eta}^2}\|Q+\bar{Q}\|\|\mathbf{x}\|_{2,\gamma}^{1/2}$.

First, by (18) and since $\bar{p}_T^T = 0$, for every $t < T$,

$$\begin{aligned}
\mathbb{E}[\|\bar{p}_t^T\|] \leq& \tilde{\gamma}\tilde{\eta}\mathbb{E}[\|\bar{p}_{t+1}^T\|] + 2\tilde{\gamma}^2\|Q+\bar{Q}\|\mathbb{E}\left[\|\bar{x}_{t+1}\|\right] \\
\leq& 2\tilde{\gamma}^2\|Q+\bar{Q}\|\sum_{i=t+1}^{T}\tilde{\gamma}^{i-t-1}\tilde{\eta}^{i-t-1}\mathbb{E}\left[\|\bar{x}_i\|\right] \\
\leq& \tilde{\gamma}^{1-t}M,
\end{aligned} \tag{21}$$

where we used the fact that $\tilde{\eta}^2 < 1$ by Assumption 1. Moreover, by equation (18), for every $0 \leq s < T_1 < \min\{T_2, T_3\}$, we have

$$\begin{aligned}
\mathbb{E}\left[\|\bar{p}_s^{T_2} - \bar{p}_s^{T_3}\|\right] \leq& \gamma\|A+\bar{A}\|\mathbb{E}\left[\|\bar{p}_{s+1}^{T_2} - \bar{p}_{s+1}^{T_3}\|\right] \\
\leq& \tilde{\gamma}^{T_1-s}\tilde{\eta}^{T_1-s}\mathbb{E}\left[\|\bar{p}_{T_1}^{T_2} - \bar{p}_{T_1}^{T_3}\|\right] \\
\leq& 2\tilde{\gamma}^{1-s}\tilde{\eta}^{T_1-s}M.
\end{aligned}$$

This concludes the proof of (20).

Claim 2. *There exist a sequence of times $(T_k)_{k\geq 0}$ and a process of random vectors $(\bar{p}_s^*)_{s\geq 0}$ satisfying: for every $s \geq 0$, $\lim_{k\to\infty}\bar{p}_s^{T_k} = \bar{p}_s^*$, $\mathbb{P}$-almost surely. Moreover*

$$\|\bar{\mathbf{p}}^*\|_{2,\gamma} < \infty. \tag{22}$$

We proceed by induction for $s = 0, 1, 2, 3, \dots$ with a diagonal argument. For time $s = 0$, by the Cauchy property of Claim 1, $(\bar{p}_0^T)_{T\geq 0}$ converges in L^1 for each of its coordinates, so for some sequence $(T_j^{(0)})_{j\geq 0}$, $(\bar{p}_0^{T_j^{(0)}})_{j\geq 0}$ converges almost surely to a random vector, say $\bar{p}_0^*$. Then, consider $s \geq 0$ and assume we have $(T_j^{(s)})_{j\geq 0}$ and $(\bar{p}_t^*)_{t\leq s}$ such that $\bar{p}_t^{T_j^{(s)}} \to \bar{p}_t^*$ as $j \to +\infty$, for all $t \leq s$. Since $(\bar{p}_{s+1}^{T_j^{(s)}})_{j\geq 0}$ is also a Cauchy sequence for convergence in L^1, there exists a sub-sequence $(T_j^{(s+1)})_{j\geq 0}$ of $(T_j^{(s)})_{j\geq 0}$ such that $(\bar{p}_{s+1}^{T_j^{(s+1)}})_{j\geq 0}$ converges almost surely to a random vector, say $\bar{p}_{s+1}^*$. We then let $T_k = T_k^{(k)}$, $k \geq 0$ to obtain the limiting process.

To prove (22), we proceed as in the proof of Proposition 2. Consider $0 \leq t < T$. We have, $\mathbb{P}-$almost surely,

$$\bar{p}_t^T = \sum_{i=0}^{T-t-1} 2\gamma^{i+1}(A+\bar{A})^i(Q+\bar{Q})\mathbb{E}\Big[\bar{x}_{t+i+1}|\mathcal{F}_t\Big].$$

As in Proposition 2, we choose γ_2 such that $0 < \gamma < \gamma_2 < 1$ and $\xi_2 = \gamma_2 \|A + \bar{A}\|^2 < 1$. Let $\eta_2 := \gamma/\gamma_2 < 1$. Then, we have:

$$\begin{aligned}
\mathbb{E}\big[\|\bar{p}_t^T\|^2\big] &\le \mathbb{E}\Big[\Big(\sum_{i=0}^{T-t-1} 2\gamma^{i+1}\|A+\bar{A}\|^i \|Q+\bar{Q}\|\mathbb{E}\big[\|\bar{x}_{t+i+1}\|\,|\mathcal{F}_t\big]\Big)^2\Big] \\
&= 4\|Q+\bar{Q}\|^2\gamma^2\mathbb{E}\Big[\Big(\sum_{i=0}^{T-t-1} \eta_2^{i/2}\xi_2^{i/2}\gamma^{i/2}\mathbb{E}[\|\bar{x}_{t+i+1}\|\,|\mathcal{F}_t]\Big)^2\Big] \\
&\le 4\|Q+\bar{Q}\|^2\gamma^2\mathbb{E}\Big[\Big(\sum_{i=0}^{T-t-1}\eta_2^i\Big)\sum_{i=0}^{T-t-1}\xi_2^i\gamma^i\mathbb{E}\big[\|\bar{x}_{t+i+1}\|^2\,|\,\mathcal{F}_t\big]\Big] \\
&\le \frac{4\|Q+\bar{Q}\|^2\gamma^{1-t}}{1-\eta_2}\Big(\sum_{j=t+1}^{T}\xi_2^{j-t-1}\mathbb{E}[\gamma^j\|\bar{x}_j\|^2]\Big).
\end{aligned}$$

Hence, summing over t and interchanging the two summations, we get,

$$\|\bar{\mathbf{p}}^T\|_{2,\gamma} \le \frac{4\|Q+\bar{Q}\|^2\gamma}{(1-\eta_2)(1-\xi_2)}\|\bar{\mathbf{x}}\|_{2,\gamma}. \tag{23}$$

For $T = T_k$ with $k \ge 0$, we apply the monotone convergence theorem and the Fatou's lemma to control the limit as $k \to \infty$. We get:

$$\|\bar{\mathbf{p}}^*\|_{2,\gamma} \le \lim_{N\to\infty}\liminf_{k\to\infty}\mathbb{E}\Big[\sum_{t=0}^{N}\gamma^t\|\bar{p}_t^{T_k}\|^2\Big] \le \frac{4\|Q+\bar{Q}\|^2\gamma}{(1-\eta_2)(1-\xi_2)}\|\bar{\mathbf{x}}\|_{2,\gamma}.$$

Proceeding analogously, we can show:

Claim 3. *There exist a subsequence $(\hat{T}_k)_{k\ge0}$ of $(T_k)_{k\ge0}$ and a process $(q_t^*)_{t\ge0}$ satisfying: for every $t \ge 0$, $\lim_{k\to\infty} p_t^{\hat{T}_k} - \mathbb{E}[p_t^{\hat{T}_k}|\mathcal{F}^0] = q_t^*$, $\mathbb{P}-$almost surely. Moreover,*

$$\|\mathbf{q}^*\|_{2,\gamma} < \infty. \tag{24}$$

Finally, we obtain an adjoint process of $\mathbf{x}$ as follows.

Claim 4. *The process $(p_t^{**})_{t\ge0}$ defined by $p_t^{**} := q_t^* + \bar{p}_t^*$, $t \ge 0$, is an adjoint process corresponding to $\mathbf{x}$ controlled by $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$.*

First, by (22) and (24), we obtain that $(p_t^{**})_{t\ge0}$ satisfies the transversality condition (16). Second, form Claim 1, we know that for every $t \ge 0$, the process of random vectors $(\bar{p}_t^T)_{T\ge t}$ converges in L^1 with the norm $\|\cdot\|$, so by Jensen's inequality, the process of conditional expectations $\big(\mathbb{E}[(A+\bar{A})^\top \bar{p}_{t+1}^{T_k} + 2(Q+\bar{Q})^\top \bar{x}_{t+1}\,\big|\,\mathcal{F}_t]\big)_{k\ge0}$ converges in L^1 when $k \to \infty$. Thus, by equation (18) and uniqueness of the limit, we obtain:

$$\bar{p}_t^* = \gamma\mathbb{E}[(A+\bar{A})^\top\bar{p}_{t+1}^* + 2(Q+\bar{Q})\bar{x}_{t+1}|\mathcal{F}_t], \quad \mathbb{P}-a.s.. \tag{25}$$

Similarly, for every $t \ge 0$, we also obtain:

$$q_t^* = \gamma\mathbb{E}\big[A^\top q_{t+1}^* + 2Q(x_{t+1} - \bar{x}_{t+1})|\mathcal{F}_t\big], \quad \mathbb{P}-a.s.\,, \tag{26}$$

and $\mathbb{P}-$almost surely, $\mathbb{E}[\bar{p}_t^*\,|\,\mathcal{F}^0] = \bar{p}_t^*$ and $\mathbb{E}[q_t^*\,|\,\mathcal{F}^0] = 0$. Adding equations (25) and (26) and using the fact that $\mathbb{E}[p_t^{**}\,|\,\mathcal{F}^0] = \bar{p}_t^*$, we conclude that the process $(p_t^{**})_{t\ge0}$ is an adjoint process of $\mathbf{x}$. □

Using the adjoint process, we express the derivative of J as follows.

Lemma 11. *The Gateaux derivative of J at $(\mathbf{u}_1, \mathbf{u}_2)$ in the direction $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \in \mathcal{U} \times \mathcal{U}$ exists and is given by*

$$DJ(\mathbf{u}_1, \mathbf{u}_2)(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \big(p_t^\top B_1 + 2u_{1,t}^\top R_1 + \bar{p}_t^\top \bar{B}_1 + 2\bar{u}_{1,t}^\top \bar{R}_1\big)\beta_{1,t}\right] + \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \big(p_t^\top B_2 - 2u_{2,t}^\top R_2 + \bar{p}_t^\top \bar{B}_2 - 2\bar{u}_{2,t}^\top \bar{R}_2\big)\beta_{2,t}\right]. \tag{27}$$

where $\mathbf{p} = (p_t)_{t\geq 0}$ is the adjoint process corresponding to the state process $\mathbf{x}$ controlled by $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}_{ad}^{open}$.

Proof. We start by computing the difference between the values of J evaluated on two pairs of controls. Let $\mathbf{u}_i = (u_{i,t})_{t=0,1,\dots}$ and $\mathbf{u}_i' = (u_{i,t}')_{t=0,1,\dots}$, $i = 1, 2$ be two pairs of admissible control processes and let us denote by x_t and x_t' the corresponding states of the system at time t, as given by the state equation (3) with the same initial point and the same realizations of the noise sequences $(\epsilon_t^0)_{t=0,1,\dots}$ and $(\epsilon_t^1)_{t=0,1,\dots}$. Note that as a consequence, $x_{t+1} - x_{t+1}'$ can be expressed as:

$$x_t - x_t' + \check{A}(x_t - x_t') + \bar{A}(\bar{x}_t - \bar{x}_t') + \sum_{i=1,2}\left[B_i(u_{i,t} - u_{i,t}') + \bar{B}_i(\bar{u}_{i,t} - \bar{u}_{i,t}')\right],$$

which shows that $x_{t+1} - x_{t+1}'$ is in fact $\mathcal{F}_t$-measurable. As before, we use the convenient notations $\zeta_t = (x_t, \bar{x}_t, u_{1,t}, \bar{u}_{1,t}, u_{2,t}, \bar{u}_{2,t})$ and $\zeta_t' = (x_t', \bar{x}_t', u_{1,t}', \bar{u}_{1,t}', u_{2,t}', \bar{u}_{2,t}')$. In order to estimate $J(\mathbf{u}_1', \mathbf{u}_2') - J(\mathbf{u}_1, \mathbf{u}_2)$ we first notice that, if $\mathbf{p} = (p_t)_{t=0,1,\dots}$ is the adjoint process of $(x_t)_{t\geq 0}$, we get:

$$\begin{aligned}\sum_{t\geq 0}\mathbb{E}[\gamma^t(x_t' - x_t)\cdot p_t] &\leq \liminf_{N\to\infty}\mathbb{E}\Big[\sum_{t=0}^{N}\gamma^t\|x_t' - x_t\|^2\Big]^{1/2}\mathbb{E}\Big[\sum_{t=0}^{N}\gamma^t\|p_t\|^2\Big]^{1/2}\\ &\leq \Big(2\mathbb{E}\Big[\sum_{t=0}^{\infty}\gamma^t\|x_t'\|^2\Big] + 2\mathbb{E}\Big[\sum_{t=0}^{\infty}\gamma^t\|x_t\|^2\Big]\Big)^{1/2}\mathbb{E}\Big[\sum_{t=0}^{\infty}\gamma^t\|p_t\|^2\Big]^{1/2},\end{aligned}$$

which is finite by the admissibility of $\mathbf{u}$ and the transversality condition (16) of $\mathbf{p}$.

Recalling the definition of $b(\zeta_t)$ in (11), by the admissible conditions of the state process $\mathbf{x}$ and the pair of control processes$(\mathbf{u}_1, \mathbf{u}_2)$, we get

$$\sum_{t\geq 0}\mathbb{E}\big[\gamma^t\big(b(\zeta_t') - b(\zeta_t)\big)\cdot p_t\big)\big] < \infty.$$

As a consequence, we have:

$$\begin{aligned}&\mathbb{E}\Big[\sum_{t=0}^{\infty}\gamma^t[c(\zeta_t') - c(\zeta_t)]\Big]\\ &= \sum_{t=0}^{\infty}\mathbb{E}\Big[\gamma^t\big(h(\zeta_t', p_t) - h(\zeta_t, p_t)\big)\Big] - \sum_{t=0}^{\infty}\mathbb{E}\Big[\gamma^t[(x_{t+1}' - x_{t+1}) - (x_t' - x_t)]\cdot p_t\\ &\qquad\qquad + \delta\sum_{t=0}^{\infty}\gamma^t(x_t' - x_t)\cdot p_t\Big]\\ &= \sum_{t=0}^{\infty}\mathbb{E}\Big[\gamma^t\big(h(\zeta_t', p_t) - h(\zeta_t, p_t)\big)\Big] + \sum_{t=0}^{\infty}\mathbb{E}\Big[\gamma^t(x_{t+1}' - x_{t+1})\cdot(p_{t+1} - p_t)\Big],\end{aligned} \tag{28}$$

where we used the bounded convergence theorem for the first equality, and we used the facts that $\delta = (1-\gamma)/\gamma$ and $p_0 = 0$ for the last equality.

We now turn to computing the Gateaux derivative of J. Let $\mathbf{u}_i, \boldsymbol{\beta}_i$, $i = 1, 2$, as in the statement. To alleviate the notation, we denote

$$V_t = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left(x_t^{\mathbf{u}_1 + \epsilon \boldsymbol{\beta}_1, \mathbf{u}_2 + \epsilon \boldsymbol{\beta}_2} - x_t^{\mathbf{u}_1, \mathbf{u}_2} \right),$$

where $x^{\mathbf{u}_1 + \epsilon \boldsymbol{\beta}_1, \mathbf{u}_2 + \epsilon \boldsymbol{\beta}_2}$ is the state process controlled by $(\mathbf{u}_1 + \epsilon \boldsymbol{\beta}_1, \mathbf{u}_2 + \epsilon \boldsymbol{\beta}_2) \in \mathcal{U}_{ad}^{open}$, and $x^{\mathbf{u}_1, \mathbf{u}_2}$ is the state process controlled by $(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}_{ad}^{open}$. By linearity of the state dynamics, we have $V_t = \left(x_t^{\mathbf{u}_1 + \epsilon \boldsymbol{\beta}_1, \mathbf{u}_2 + \epsilon \boldsymbol{\beta}_2} - x_t^{\mathbf{u}_1, \mathbf{u}_2} \right) / \epsilon$ for every $\epsilon > 0$ and every $t \geq 0$. Let $(\mathbf{u}_1', \mathbf{u}_2') = (\mathbf{u}_1 + \epsilon \boldsymbol{\beta}_1, \mathbf{u}_2 + \epsilon \boldsymbol{\beta}_2)$.

We then compute, using the expressions of the partial derivatives of h already computed in the proof of Lemma 8:

$$\begin{aligned}
&DJ(\mathbf{u}_1, \mathbf{u}_2)(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[J(\mathbf{u}_1 + \epsilon \boldsymbol{\beta}_1, \mathbf{u}_2 + \epsilon \boldsymbol{\beta}_2) - J(\mathbf{u}_1, \mathbf{u}_2) \right] \\
&= \sum_{t=0}^{\infty} \mathbb{E} \left[\gamma^t V_{t+1} \cdot (p_{t+1} - p_t) \right] + \sum_{t=0}^{\infty} \gamma^t \mathbb{E} \Big[\partial_x h(\zeta_t, p_t) V_t + \partial_{\bar{x}} h(\zeta_t, p_t) \bar{V}_t \\
&\qquad + \partial_{u_1} h(\zeta_t, p_t) \beta_{1,t} + \partial_{\bar{u}_1} h(\zeta_t, p_t) \bar{\beta}_{1,t} + \partial_{u_2} h(\zeta_t, p_t) \beta_{2,t} + \partial_{\bar{u}_2} h(\zeta_t, p_t) \bar{\beta}_{2,t} \Big] \\
&= \sum_{t=0}^{\infty} \gamma^t \mathbb{E} \Big[\underbrace{V_{t+1} \cdot (p_{t+1} - p_t)}_{(0)} \\
&\quad + \underbrace{\left(p_t^\top (\check{A} - \delta I_d) + 2(x_t - \bar{x}_t)^\top Q \right) V_t}_{(i)} + \underbrace{\left(p_t^\top \bar{A} + 2(\bar{x}_t - x_t)^\top Q + 2 \bar{x}_t^\top (Q + \bar{Q}) \right) \bar{V}_t}_{(ii)} \\
&\quad + \underbrace{\left(p_t^\top B_1 + 2(u_{1,t} - \bar{u}_{1,t})^\top R_1 \right) \beta_{1,t}}_{(iii)_1} + \underbrace{\left(p_t^\top B_2 - 2(u_{2,t} - \bar{u}_{2,t})^\top R_2 \right) \beta_{2,t}}_{(iii)_2} \\
&\quad + \underbrace{\left(p_t^\top \bar{B}_1 + 2(\bar{u}_{1,t} - u_{1,t})^\top R_1 + 2 \bar{u}_{1,t}^\top (R_1 + \bar{R}_1) \right) \bar{\beta}_{1,t}}_{(iv)_1} \\
&\quad + \underbrace{\left(p_t^\top \bar{B}_2 - 2(\bar{u}_{2,t} - u_{2,t})^\top R_2 - 2 \bar{u}_{2,t}^\top (R_2 + \bar{R}_2) \right) \bar{\beta}_{2,t}}_{(iv)_2} \Big].
\end{aligned} \tag{29}$$

We now use Fubini's theorem to compute two of the six terms above. Recall that $\bar{V}_t = \mathbb{E}[V_t | \mathcal{F}^0]$, which we choose to express in the form

$$\bar{V}_t = \tilde{\mathbb{E}}[\tilde{V}_t | \mathcal{F}^0] = \int_{\tilde{\Omega}^1} \tilde{V}_t(\omega^0, \tilde{\omega}^1) \tilde{\mathbb{P}}^1(d\tilde{\omega}^1),$$

where $(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1)$ is an identical copy of $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ and the probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ is defined as $\tilde{\Omega} = \Omega^0 \times \tilde{\Omega}^1$, $\tilde{\mathcal{F}} = \mathcal{F}^0 \times \tilde{\mathcal{F}}^1$,and $\tilde{\mathbb{P}} = \mathbb{P}^0 \times \tilde{\mathbb{P}}^1$. For the sake of ease of notation, we introduce yet another notation for the conditional expectations: we shall denote by $\mathbb{E}_{\mathcal{F}^0}$ and $\tilde{\mathbb{E}}_{\mathcal{F}^0}$ the conditional expectations usually denoted by $\mathbb{E}[\cdot\,|\mathcal{F}^0]$ and $\tilde{\mathbb{E}}[\cdot\,|\mathcal{F}^0]$ respectively. With this new notation $\bar{x}_t = \mathbb{E}_{\mathcal{F}^0}[x_t] = \tilde{\mathbb{E}}_{\mathcal{F}^0}[\tilde{x}_t] = \bar{\tilde{x}}_t$ and similarly for the other random variables. Consequently:

$$\begin{aligned}\mathbb{E}\sum_{t=0}^{\infty}\gamma^t(ii) &= \mathbb{E}\mathbb{E}_{\mathcal{F}^0}\sum_{t=0}^{\infty}\gamma^t\big(p_t^\top\bar{A}+2(\bar{x}_t-x_t)^\top Q+2\bar{x}_t^\top(Q+\bar{Q})\big)\bar{V}_t\\ &= \mathbb{E}\mathbb{E}_{\mathcal{F}^0}\tilde{\mathbb{E}}_{\mathcal{F}^0}\sum_{t=0}^{\infty}\gamma^t\big(p_t^\top\bar{A}+2(\bar{x}_t-x_t)^\top Q+2\bar{x}_t^\top(Q+\bar{Q})\big)\tilde{V}_t\\ &= \mathbb{E}\sum_{t=0}^{\infty}\gamma^t\big(\bar{p}_t^\top\bar{A}+2\bar{x}_t^\top(Q+\bar{Q})\big)V_t,\end{aligned}$$

where we used Fubini's theorem for the last equality. So:

$$\mathbb{E}\sum_{t=0}^{\infty}\gamma^t[(i)+(ii)] = \mathbb{E}\sum_{t=0}^{\infty}\gamma^t\big(p_t^\top(\check{A}-\delta I_d)+2x_t^\top Q+\bar{p}_t^\top\bar{A}+2\bar{x}_t^\top\bar{Q}\big)V_t. \tag{30}$$

As a consequence, $\sum_{t=0}^{\infty}\gamma^t\mathbb{E}\big[(0)+(i)+(ii)\big]=0$, because of (30) and the definition (15) of the adjoint process.

Furthermore, using Fubini's theorem on an identical copy of $\mathbf{u}_1$ and $\boldsymbol{\beta}_1$ we get:

$$\mathbb{E}\sum_{t=0}^{\infty}\gamma^t[(iii)_1+(iv)_1] = \mathbb{E}\sum_{t=0}^{\infty}\gamma^t\big(p_t^\top B_1+2u_{1,t}^\top R_1+\bar{p}_t^\top\bar{B}_1+2\bar{u}_{1,t}^\top\bar{R}_1\big)\beta_{1,t}, \tag{31}$$

and likewise for $\mathbf{u}_2,\boldsymbol{\beta}_2$. □

We are now in a position to prove the following condition for optimality:

Proposition 12 (Pontryagin's maximum principle, necessary condition)**.** *Assuming that Assumption 1 holds, if $\mathbf{u}_i=(u_{i,t})_{t=0,1,\dots}$, $i=1,2$ is a pair of admissible control processes such that it is an open-loop saddle point for the zero-sum game and $\mathbf{p}=(p_t)_{t=0,1,\dots}$ is the corresponding adjoint process, then it holds*

$$\begin{cases} B_1^\top p_t+2R_1u_{1,t}+\bar{B}_1^\top\bar{p}_t+2\bar{R}_1\bar{u}_{1,t}=0\\ B_2^\top p_t-2R_2u_{2,t}+\bar{B}_2^\top\bar{p}_t-2\bar{R}_2\bar{u}_{2,t}=0\end{cases} \tag{32}$$

for all $t\geq 0$, $\mathbb{P}$-almost surely.

Proof. By Lemma 11, for any pair of processes $(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)\in\mathcal{U}\times\mathcal{U}$ we have the

$$\begin{aligned}DJ(\mathbf{u}_1,\mathbf{u}_2)(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2) = \mathbb{E}&\left[\sum_{t=0}^{\infty}\gamma^t\big(p_t^\top B_1+2u_{1,t}^\top R_1+\bar{p}_t^\top\bar{B}_1+2\bar{u}_{1,t}^\top\bar{R}_1\big)\beta_{1,t}\right]\\ &+\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t\big(p_t^\top B_2-2u_{2,t}^\top R_2+\bar{p}_t^\top\bar{B}_2-2\bar{u}_{2,t}^\top\bar{R}_2\big)\beta_{2,t}\right].\end{aligned}$$

Since $(\mathbf{u}_1,\mathbf{u}_2)\in\mathcal{U}_{ad}^{open}$ is an open-loop saddle point for the zero-sum game, then for every $\mathbf{u}_1'\in\mathcal{U}$ and $\mathbf{u}_2'\in\mathcal{U}$, we have

$$J(\mathbf{u}_1,\mathbf{u}_2')\leq J(\mathbf{u}_1,\mathbf{u}_2)\leq J(\mathbf{u}_1',\mathbf{u}_2).$$

Denote the state processes in the above inequalities by $\boldsymbol{X}^{\mathbf{u}_1,\mathbf{u}_2'},\boldsymbol{X}^{\mathbf{u}_1,\mathbf{u}_2},\boldsymbol{X}^{\mathbf{u}_1',\mathbf{u}_2}$, which are all L^2-discounted globally integrable according to Proposition 2.

If we choose $\boldsymbol{\beta}_2=0$, then for every $\boldsymbol{\beta}_1\in\mathcal{U}$,

$$DJ(\mathbf{u}_1,\mathbf{u}_2)(\boldsymbol{\beta}_1,0)=\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[J(\mathbf{u}_1+\epsilon\boldsymbol{\beta}_1,\mathbf{u}_2)-J(\mathbf{u}_1,\mathbf{u}_2)\right]\geq 0$$

which implies,

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t\big(p_t^\top B_1+2u_{1,t}^\top R_1+\bar p_t^\top \bar B_1+2\bar u_{1,t}^\top \bar R_1\big)\beta_{1,t}\right]\geq 0.$$

Thus, the corresponding adjoint process $\mathbf{p}$ satisfies: $\mathbb{P}-$almost surely,

$$B_1^\top p_t+2R_1u_{1,t}+\bar B_1^\top \bar p_t+2\bar R_1\bar u_{1,t}=0,\qquad t\geq 0.$$

Similarly, $B_2^\top p_t-2R_2u_{2,t}+\bar B_2^\top \bar p_t-2\bar R_2\bar u_{2,t}=0$ for all $t\geq 0$, $\mathbb{P}$-almost surely. □

4.3. **Identification of the equilibrium.** Let us introduce the notations

$$\begin{cases}\Gamma_i=(-1)^i\frac{1}{2}R_i^{-1}B_i^\top, \qquad \Xi_i=(-1)^i\frac{1}{2}R_i^{-1}\big[\bar B_i^\top-\bar R_i(R_i+\bar R_i)^{-1}(B_i+\bar B_i)^\top\big],\\ \Lambda_i=\Gamma_i+\Xi_i=(-1)^i\frac{1}{2}(R_i+\bar R_i)^{-1}(B_i+\bar B_i)^\top, i=1,2.\end{cases}\tag{33}$$

We then consider the following Riccati equations:

$$\gamma[A^\top P+2Q]\big(A+\big(B_1\Gamma_1+B_2\Gamma_2\big)P\big)=P,\tag{34}$$

and

$$\gamma\big[(A^\top+\bar A^\top)\bar P+2(Q+\bar Q)\big]\left[(A+\bar A)+\sum_{i=1,2}(B_i+\bar B_i)\Lambda_i\bar P\right]=\bar P.\tag{35}$$

We shall assume that there exist solutions P and $\bar P$ in $\mathcal{S}^d$ to these equations. This can be proved under suitable conditions, for example, by contraction arguments when some coefficients are small enough. We also discuss in section 6 a way to construct P and $\bar P$ with the help of other Algebraic Riccati equations (ARE for short).

We now rewrite the equilibrium condition (32) in Proposition 12. The process $\mathbf{u}=(\mathbf{u}_1,\mathbf{u}_2)$ is an OLSP and the process $\mathbf{p}$ is the corresponding adjoint process. Taking conditional expectations $\mathbb{E}_{\mathcal{F}^0}$ in the first equation, we get:

$$(B_1+\bar B_1)^\top \bar p_t+2(R_1+\bar R_1)\bar u_{1,t}=0$$

from which we derive:

$$\bar u_{1,t}=-\frac{1}{2}(R_1+\bar R_1)^{-1}(B_1+\bar B_1)^\top \bar p_t.\tag{36}$$

Plugging this expression back into the first equation of (32) we deduce:

$$u_{1,t}=\Gamma_1p_t+\Xi_1\bar p_t,\qquad\text{and}\qquad \bar u_{1,t}=\Lambda_1\bar p_t\tag{37}$$

for Γ_1,Ξ_1,Λ_1 introduced in (33). Similarly, we find

$$u_{2,t}=\Gamma_2p_t+\Xi_2\bar p_t,\qquad\text{and}\qquad \bar u_{2,t}=\Lambda_2\bar p_t.\tag{38}$$

Proposition 13. *Assume there exist solutions P and $\bar P$ of (34)–(35), and that:*

$$\gamma\left\|A+\sum_{i=1,2}B_i\Gamma_iP\right\|^2<1,\qquad \gamma\left\|(A+\bar A)+\sum_{i=1,2}(B_i+\bar B_i)\Lambda_i\bar P\right\|^2<1.\tag{39}$$

Let $\mathbf{x}$ be the process defined by: $x_0=\epsilon_0^0+\epsilon_0^1$ and for $t\geq 0$,

$$x_{t+1}=Ax_t+\bar A\bar x_t+\sum_{i=1,2}\big[B_i(\Gamma_iP(x_t-\bar x_t)+\Lambda_i\bar P\bar x_t)+\bar B_i\Lambda_i\bar P\bar x_t\big]+\epsilon_{t+1}^0+\epsilon_{t+1}^1.\tag{40}$$

Let $\mathbf{p}$ be the process defined by:

$$p_t=P(x_t-\bar x_t)+\bar P\bar x_t,\qquad t\geq 0.\tag{41}$$

Let $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$ *be the process defined by:*

$$u_{i,t} = \Gamma_i P(x_t - \bar{x}_t) + \Lambda_i \bar{P} \bar{x}_t, \qquad i = 1, 2, \quad t \geq 0. \tag{42}$$

Then $\mathbf{u}$ *is an admissible pair of controls and* $\mathbf{p}$ *is the associated adjoint process.*

Condition (39) can be satisfied for instance by assuming that the coefficients of the problem are small enough.

Proof. We first check the admissibility. We note that the dynamics (40) amounts to (3) with control pair $\mathbf{u}$ defined by (42). Moreover, $\bar{\mathbf{x}}$ satisfies:

$$\bar{x}_{t+1} = \tilde{\Lambda} \bar{x}_t + \epsilon^0_{t+1} = \tilde{\Lambda}^{t+1} \bar{x}_0 + \sum_{j=1}^{t+1} \tilde{\Lambda}^{t+1-j} \epsilon^0_j,$$

where $\tilde{\Lambda} = \Big[(A + \bar{A}) + \sum_{i=1,2} (B_i + \bar{B}_i) \Lambda_i \bar{P}\Big]$. Hence

$$\|\bar{\mathbf{x}}\|_{2,\gamma} \leq \mathbb{E}\Big[\sum_{t \geq 0} \gamma^t \Big(\|\tilde{\Lambda}\|^{2t} \|\bar{x}_0\|^2 + \|\sum_{j=1}^{t} \tilde{\Lambda}^{t-j} \epsilon^0_j\|^2 \Big)\Big],$$

which is finite by (39) and $\epsilon^0_t, t \geq 0$, have a finite second moment.

Similarly, with $\tilde{\Gamma} = A + \sum_{i=1,2} B_i \Gamma_i P$,

$$\|\mathbf{x} - \bar{\mathbf{x}}\|_{2,\gamma} \leq \mathbb{E}\Big[\sum_{t \geq 0} \gamma^t \Big(\|\tilde{\Gamma}\|^{2t} \|x_0 - \bar{x}_0\|^2 + \|\sum_{j=1}^{t} \tilde{\Gamma}^{t-j} \epsilon^1_j\|^2 \Big)\Big],$$

which is finite. Hence the control pair $\mathbf{u}$ is admissible and transversality condition (16) is satisfied.

Furthermore, we have:

$$p_{t+1} - p_t = -\gamma\big[(\check{A}^\top - \delta I_d) p_{t+1} + 2Q x_{t+1} + \bar{A}^\top \bar{p}_{t+1} + 2\bar{Q} \bar{x}_{t+1}\big] + Z^0_{t+1} \epsilon^0_{t+1} + Z^1_{t+1} \epsilon^1_{t+1}, \tag{43}$$

where the processes $\boldsymbol{Z}^0$ and $\boldsymbol{Z}^1$ are deterministic and independent of time, and defined by:

$$Z^0_t = \gamma[(A^\top + \bar{A}^\top)\bar{P} + 2(Q + \bar{Q})], \qquad \text{and} \qquad Z^1_t = \gamma[A^\top P + 2Q].$$

Hence $\mathbf{p}$ is indeed the adjoint process associated to $\mathbf{u}$. □

4.4. **A convexity-concavity sufficient condition.** Consider two deterministic processes $\mathbf{V}_1 = (V_{1,t})_{t \geq 0}$ and $\mathbf{V}_2 = (V_{2,t})_{t \geq 0}$ following the dynamics

$$\begin{cases} V_{1,t+1} = A V_{1,t} + \bar{A} \bar{V}_{1,t} + B_1 \beta_{1,t} + \bar{B}_1 \bar{\beta}_{1,t}, & V_{1,t=0} = 0, \qquad (44a) \\ V_{2,t+1} = A V_{2,t} + \bar{A} \bar{V}_{2,t} + B_2 \beta_{2,t} + \bar{B}_2 \bar{\beta}_{2,t}, & V_{2,t=0} = 0, \qquad (44b) \end{cases}$$

where $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \in \mathcal{U} \times \mathcal{U}$ are two L^2-integrable control processes. Under Assumption 1, by Lemma 3, we have $\mathbf{V}_1 \in \mathcal{X}$ and $\mathbf{V}_2 \in \mathcal{X}$.

Proposition 14 (Pontryagin's maximum principle, sufficient condition)**.** *We assume the following conditions:*

1. *There exists a state process* $\mathbf{x} = (x_t)_{t\geq 0}$ *and* $\mathbf{p} = (p_t)_{t\geq 0}$ *such that* $\mathbf{x}, \mathbf{p}$ *are* $(\mathcal{F}_t)_{t\geq 0}$*-adapted,* L^2-*discounted globally integrable, and they satisfy the forward-backward system of equations: for every* $t \geq 0$*,*

$$\begin{cases} x_{t+1} &= Ax_t + \bar{A}\bar{x}_t + (B_1\Gamma_1 + B_2\Gamma_2)p_t \\ &\quad + \Big((B_1+\bar{B}_1)\Lambda_1 + (B_2+\bar{B}_2)\Lambda_2 - B_1\Gamma_1 - B_2\Gamma_2\Big)\bar{p}_t + \epsilon^0_{t+1} + \epsilon^1_{t+1}, \\ p_t &= \gamma\left(A^\top p_{t+1} + 2Qx_{t+1} + \bar{A}^\top \bar{p}_{t+1} + 2\bar{Q}\bar{x}_{t+1}\right) + Z^0_{t+1}\epsilon^0_{t+1} + Z^1_{t+1}\epsilon^1_{t+1} \end{cases} \tag{45}$$

with initial values $x_0 = \epsilon^0_0 + \epsilon^1_0$ *and for some* $(\mathcal{F}_t)_{t\geq 0}$*-predictable processes* $(Z^0_t, Z^1_t)_{t\geq 1}$ *satisfying* $Z^0_0 = Z^1_0 = 0$*.*

2. *For any control processes* $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \in \mathcal{U}\times\mathcal{U}$*, we have the following convexity-concavity condition for the zero-sum game:*

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t\left(V_{1,t}^\top Q V_{1,t} + \bar{V}_{1,t}^\top \bar{Q}\bar{V}_{1,t} + \beta_{1,t}^\top R_1\beta_{1,t} + \bar{\beta}_{1,t}^\top \bar{R}_1\bar{\beta}_{1,t}\right)\right] \geq 0 \tag{46}$$

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t\left(V_{2,t}^\top Q V_{2,t} + \bar{V}_{2,t}^\top \bar{Q}\bar{V}_{2,t} - \beta_{2,t}^\top R_2\beta_{2,t} - \bar{\beta}_{2,t}^\top \bar{R}_2\bar{\beta}_{2,t}\right)\right] \leq 0 \tag{47}$$

where the processes $(\mathbf{V}_1, \mathbf{V}_2) \in \mathcal{X}\times\mathcal{X}$ *follows the dynamics* (44a)–(44b)*.*

Then, the pair of control processes $(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}_{ad}^{open}$ *given by:*

$$u_{i,t} = \Gamma_i p_t + (\Lambda_i - \Gamma_i)\bar{p}_t, \qquad i = 1, 2 \tag{48}$$

is an OLSP for the zero-sum game. Moreover, $(\mathbf{u}_1, \mathbf{u}_2)$ *satisfies the equilibrium condition* (32) *of the Pontryagin maximum principle.*

Proof. The backward equation for process $\mathbf{p}$ implies that it satisfies the conditional expectation condition (15). We show with equations (37)–(38) that the pair of control processes $(\mathbf{u}_1, \mathbf{u}_2)$ defined by equation (48) satisfies the equilibrium condition (32). By substituting the right hand side of (48) with $(u_{1,t}, u_{2,t})$ in the forward equation for $(x_t)_{t\geq 0}$ in (45), we get that the process $\mathbf{x}$ follows dynamics (3) which is controlled exactly by $(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}_{ad}^{open}$.

Based on the proof of Lemma 11 for the Gateaux derivative of J, we write a second-order expansion for the value function J at a point $(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}_{ad}^{open}$ in the direction $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \in \mathcal{U}\times\mathcal{U}$. To alleviate the notation, we introduce a deterministic process $\mathbf{V} = (V_t)_{t\geq 0}$ following a dynamics

$$V_{t+1} = AV_t + \bar{A}\bar{V}_t + B_1\beta_1 + \bar{B}_1\bar{\beta}_1 + B_2\beta_2 + \bar{B}_2\beta_2$$

with initial value $V_0 = 0$. The linearity of the dynamics (3) shows that $V_t = (x_t^{\mathbf{u}_1+\epsilon\boldsymbol{\beta}_1, \mathbf{u}_2+\epsilon\boldsymbol{\beta}_2} - x_t^{\mathbf{u}_1,\mathbf{u}_2})/\epsilon$ for every $\epsilon > 0$ and $t \geq 0$. According to equation (28), the difference between the values of J at point $(\mathbf{u}_1 + \epsilon\boldsymbol{\beta}_1, \mathbf{u}_2 + \epsilon\boldsymbol{\beta}_2)$ and at point $(\mathbf{u}_1, \mathbf{u}_2)$ can be expressed by

$$\begin{aligned} &J(\mathbf{u}_1 + \epsilon\boldsymbol{\beta}_1, \mathbf{u}_2 + \epsilon\boldsymbol{\beta}_2) - J(\mathbf{u}_1, \mathbf{u}_2) \\ =&\epsilon\left(\sum_{t=0}^{\infty}\mathbb{E}\left[\gamma^t V_{t+1}\cdot(p_{t+1} - p_t)\right] + \sum_{t=0}^{\infty}\gamma^t\mathbb{E}\left[\nabla_\zeta h(\zeta_t, p_t)\cdot\check{\zeta}_t\right]\right) \\ &\qquad + \frac{1}{2}\epsilon^2\sum_{t=0}^{\infty}\gamma^t\mathbb{E}\Big[\nabla^2_{\zeta\zeta}h(\eta_t, p_t)\check{\zeta}_t\cdot\check{\zeta}_t\Big] \\ =&\epsilon(i) + \epsilon^2(ii), \end{aligned} \tag{49}$$

where $\check{\zeta}_t = (\zeta'_t - \zeta_t)/\epsilon = \big[V_t^\top, \bar{V}_t^\top, \beta_{1,t}^\top, \bar{\beta}_{1,t}^\top, \beta_{2,t}^\top, \bar{\beta}_{2,t}^\top\big]^\top \in \mathbb{R}^{2d\times 4\ell}$ and $\eta_t = (1-\lambda_t)\zeta'_t + \lambda_t\zeta_t \in \mathbb{R}^{2d+4\ell}$ for some $\lambda_t \in [0,1]$. Since the pair of admissible control processes $(\mathbf{u}_1, \mathbf{u}_2)$ satisfies the system of equations (32) at every time $t \geq 0$, then by applying Lemma 11, we have $(i) = 0$. We also notice that the Hessian of $h(\cdot)$ with respect to ζ is a constant matrix depending only on model parameters. Thus, we obtain

$$(ii) = \sum_{t=0}^{\infty} \gamma^t \mathbb{E}\Big[V_t^\top Q V_t + \bar{V}_t^\top \bar{Q} \bar{V}_t + \beta_{1,t}^\top R_1 \beta_{1,t} + \bar{\beta}_{1,t}^\top \bar{R}_1 \bar{\beta}_{1,t} - \beta_{2,t}^\top R_2 \beta_{2,t} - \bar{\beta}_{2,t}^\top \bar{R}_2 \bar{\beta}_{2,t}\Big]. \tag{50}$$

Consider a fixed control process $\mathbf{u}_2$ for player 2. For every control process $\mathbf{u}'_1 \in \mathcal{U}$, we choose $\boldsymbol{\beta}_1 = (\mathbf{u}'_1 - \mathbf{u}_1)/\epsilon \in \mathcal{U}$ and $\boldsymbol{\beta}_2 = 0$. The convexity condition (46), together with (49) and (50), yield that for every $\mathbf{u}'_1 \in \mathcal{U}$, $J(\mathbf{u}'_1, \mathbf{u}_2) \geq J(\mathbf{u}_1, \mathbf{u}_2)$. Similarly, the concavity condition (47) implies that for every $\mathbf{u}'_2 \in \mathcal{U}$, $J(\mathbf{u}_1, \mathbf{u}'_2) \leq J(\mathbf{u}_1, \mathbf{u}_2)$.

Therefore, we conclude that under the convexity-concavity condition for the two processes $(\mathbf{V}_1, \mathbf{V}_2)$, a pair of control processes $(\mathbf{u}_1, \mathbf{u}_2)$ satisfying the system of equations (32) with adjoint process $\mathbf{p}$ is an OLSP for the zero-sum game. □

Remark 15. We can see from equations (49) and (50) that the convexity-concavity condition is also a necessary condition if $(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}_{ad}^{open}$ is an OLSP for the zero-sum game.

Taking a closer look at the convexity condition (46) (resp. the concavity condition (47)), it is indeed a quadratic function of the process $\mathbf{V}_1$ (resp. $\mathbf{V}_2$) and the control $\boldsymbol{\beta}_1 \in \mathcal{U}$ (resp. $\boldsymbol{\beta}_2 \in \mathcal{U}$). So, we can apply results from the deterministic Linear-Quadratic control problems to derive a sufficient condition for the convexity-concavity condition. Let us define some new value functions $C_{V_i}(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}_i)$ and $\bar{C}_{V_i}(\bar{\boldsymbol{\beta}}_i)$ for $i = 1, 2$:

$$\begin{aligned} C_{V_i}(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}_i) = \mathbb{E}\Big[\sum_{t=0}^{\infty} \gamma^t \Big((V_{i,t} - \bar{V}_{i,t})^\top \Big((-1)^{i-1} Q \Big) (V_{i,t} - \bar{V}_{i,t}) \\ + (\beta_{i,t} - \bar{\beta}_{i,t})^\top R_i (\beta_{i,t} - \bar{\beta}_{i,t}) \Big) \Big], \\ \bar{C}_{V_i}(\bar{\boldsymbol{\beta}}_i) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \left(\bar{V}_{i,t}^\top \Big((-1)^{i-1} (Q + \bar{Q}) \Big) \bar{V}_{i,t} + \bar{\beta}_{i,t}^\top (R_i + \bar{R}_i) \bar{\beta}_{i,t} \right)\right]. \end{aligned}$$

Then the convexity-concavity condition (46)–(47) is equivalent to

$$\min_{\boldsymbol{\beta}_1 \in \mathcal{U}} C_{V_1}(\boldsymbol{\beta}_1 - \bar{\boldsymbol{\beta}}_1) + \bar{C}_{V_1}(\bar{\boldsymbol{\beta}}_1) \geq 0, \qquad \text{and} \qquad \min_{\boldsymbol{\beta}_2 \in \mathcal{U}} C_{V_2}(\boldsymbol{\beta}_2 - \bar{\boldsymbol{\beta}}_2) + \bar{C}_{V_2}(\bar{\boldsymbol{\beta}}_2) \geq 0.$$

Here, we multiply Q and $Q + \bar{Q}$ by -1 for player $i = 2$ so that the concavity condition is connected to a minimization problem. Let us assume that the following discrete Algebraic Riccati equation (DARE-i):

$$0 = (-1)^{i-1} Q - P_i + \gamma A^\top P_i A - \gamma^2 A^\top P_i B_i \Big(\gamma B_i^\top P_i B_i + R_i \Big)^{-1} B_i^\top P_i A, \tag{51}$$

admits a symmetric matrix $P_i \in \mathcal{S}^d$ as solution satisfying $\gamma B_i^\top P_i B_i + R_i \succ 0$ and $\gamma \|A - B_i K_i\|^2 < 1$ where $K_i = \gamma(\gamma B_i^\top P_i B_i + R_i)^{-1} B_i^\top P_i A$. Then, by applying the Dynamically Programming principle and the expression of optimal value function [45] starting at time $t = 1$ with an initial $V_{i,1} = B_i(\beta_{i,0} - \bar{\beta}_{i,0})$, the value $C_{V_i}(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}_i)$

can be expressed as:

$$\min_{\boldsymbol{\beta}_i} C_{V_i}(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}_i) = \min_{\beta_{i,0}=\check{\beta}_i} \mathbb{E}\left[\min_{\boldsymbol{\beta}_i:\beta_{i,0}=\check{\beta}_i} C_{V_i}(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}_i)\right]$$
$$= \min_{\check{\beta}_i} \mathbb{E}\left[(\check{\beta}_i - \bar{\check{\beta}}_i)^\top \Big(R_i + \gamma B_i^\top P_i B_i\Big)(\check{\beta}_i - \bar{\check{\beta}}_i)\right] \geq 0.$$

Existence of such a solution P_i can be guaranteed under suitable conditions, see [52, Theorem 3.1]. Moreover, for every given random variable $\check{\beta}_i$, the value given by $\min_{\boldsymbol{\beta}_i:\beta_{i,0}=\check{\beta}_i} C_{V_i}(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}_i)$ is attained at $\beta_{i,t} - \bar{\beta}_{i,t} = -K_i(V_{i,t} - \bar{V}_{i,t})$ for $t \geq 1$.

Similarly, if the discrete Algebraic Riccati equation for $i = 1, 2$ (DARE-MF-i):

$$0 = (-1)^{i-1}(Q + \bar{Q}) - \bar{P}_i + \gamma(A + \bar{A})^\top \bar{P}_i(A + \bar{A}) \tag{52}$$
$$- \gamma^2(A + \bar{A})^\top \bar{P}_i(B_i + \bar{B}_i)\Big(\gamma(B_i + \bar{B}_i)^\top \bar{P}_i(B_i + \bar{B}_i)$$
$$+ (R_i + \bar{R}_i)\Big)^{-1}(B_i + \bar{B}_i)^\top \bar{P}_i(A + \bar{A}),$$

has a solution $\bar{P}_i$ such that $\gamma(B_i + \bar{B}_i)^\top \bar{P}_i(B_i + \bar{B}_i) + (R_i + \bar{R}_i) \succ 0$ and $\gamma\|A + \bar{A} - (B_i + \bar{B}_i)L_i\|^2 < 1$ where $L_i = \gamma\Big(\gamma(B_i + \bar{B}_i)^\top \bar{P}_i(B_i + \bar{B}_i) + (R_i + \bar{R}_i)\Big)^{-1}(B_i + \bar{B}_i)^\top \bar{P}_i(A + \bar{A})$, then the value function $\bar{C}_{V_i}(\bar{\boldsymbol{\beta}}_i)$ can be expressed as

$$\min_{\bar{\boldsymbol{\beta}}_i} \bar{C}_{V_i}(\bar{\boldsymbol{\beta}}_i) = \min_{\check{\bar{\beta}}_i} \mathbb{E}\left[(\check{\bar{\beta}}_i)^\top \Big((R_i + \bar{R}_i) + \gamma(B_i + \bar{B}_i)^\top \bar{P}_i(B_i + \bar{B}_i)\Big)\check{\bar{\beta}}_i\right] \geq 0.$$

Furthermore, for every given random variable $\check{\bar{\beta}}_i$, the value $\min_{\bar{\boldsymbol{\beta}}_i:\bar{\beta}_{i,0}=\check{\bar{\beta}}_i} \bar{C}_{V_i}(\bar{\boldsymbol{\beta}}_i)$ is attained at $\bar{\beta}_{i,t} = -L_i\bar{V}_{i,t}$ for $t \geq 1$.

We have directly the following sufficient condition for the convexity-concavity condition.

Lemma 16. *If the four discrete Algebraic Riccati equations (DARE-i) and (DARE-MF-i) for $i = 1, 2$, i.e., (51) and (52), have solutions $(P_i, \bar{P}_i)$ such that*

$$\gamma B_i^\top P_i B_i + R_i \succ 0, \qquad \gamma(B_i + \bar{B}_i)^\top \bar{P}_i(B_i + \bar{B}_i) + (R_i + \bar{R}_i) \succ 0, \tag{53}$$

then the convexity-concavity condition (46)–(47) for the process $\mathbf{V}_1$ and $\mathbf{V}_2$ holds.

Together with the sufficient condition of the Pontryagin maximum principle (Proposition 14), we have

Corollary 17. *Let $(P_1, P_2, \bar{P}_1, \bar{P}_2)$ be solutions to the four discrete Algebraic Riccati equations in Lemma 16, and they satisfy conditions (53), then the pair of control processes $(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}_{ad}^{open}$ defined in Proposition 14 by (48) is an OLSP for the zero-sum game.*

To conclude this section, we propose a sufficient condition for the existence of $(P_1, P_2, \bar{P}_1, \bar{P}_2)$ in Corollary 17. We say that (A, B_i) is $\gamma-$stabilizable if there exists a matrix $K \in \mathbb{R}^{\ell \times d}$ such that all eigenvalues of $\gamma^{1/2}(A - B_iK)$ in the complex plan lie inside the unit circle, i.e. $\gamma\|A - B_iK\|^2 < 1$.

Similarly, for the existence of $\bar{P}_i$ to (DARE-MF-i) satisfying condition (53), we can define $\bar{\mathcal{R}}_i(\eta)$ as the right hand side of (52) and consider the set

$$\bar{\mathcal{D}}_i := \left\{\bar{\eta} \in \mathcal{S}^d \mid \gamma(B_i + \bar{B}_i)^\top \bar{\eta}(B_i + \bar{B}_i) + (R_i + \bar{R}_i) \succ 0,\ \bar{\mathcal{R}}_i(\bar{\eta}) \succeq 0\right\}.$$

We notice that if $\gamma\|A\|^2 < 1$, (A, B_i) is $\gamma-$stabilizable, for $i = 1, 2$. So, under Assumption 1, $(A, B_1), (A, B_2), (\tilde{A}, \tilde{B}_1)$ and $(\tilde{A}, \tilde{B}_2)$ are all $\gamma-$stabilizable.

Corollary 18. *Assume Assumption 1, and assume that* $Q, \bar{Q} \in \mathcal{S}^d$, $R_i, R_i + \bar{R}_i \succ 0$ *for* $i = 1, 2$. *If* $\mathcal{D}_1, \mathcal{D}_2, \bar{\mathcal{D}}_1, \bar{\mathcal{D}}_2 \neq \emptyset$, *and if the forward-backward system of equations* (45) *holds for* $\mathbf{x}$ *and* $\mathbf{p}$, *then there exists an OLSP for the zero-sum game.*

5. Closed-loop information structure.

In this section, we turn our attention to closed-loop controls, that is, controls which are functions of the state and the conditional mean. We will in fact focus on a specific class of such functions.

5.1. Admissible set of controls.

We start by defining the class of functions we will consider in the rest of this section.

Definition 19. For $i = 1, 2$, a closed-loop feedback strategy (or policy) for player i is a function $v_{\theta_i} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^\ell$, $(x, \bar{x}) \mapsto (-1)^i K_i(x - \bar{x}) + (-1)^i L_i \bar{x}$ parameterized by a tuple $\theta_i = (K_i, L_i)$ where K_i and L_i are (deterministic) matrices in $\mathbb{R}^{\ell \times d}$. A pair of policies given by parameters $(\theta_1, \theta_2) \in (\mathbb{R}^{\ell \times d})^2 \times (\mathbb{R}^{\ell \times d})^2$ for the two players is called a closed-loop feedback policy profile.

For simplicity, in the sequel, we will use interchangeably the terms strategy, policy and parameters. In other words, for any θ, we identify the parameters θ with the induced closed-loop policy v_θ.

We consider the following set of admissible policies in the closed-loop setting:

$$\Theta_{ad}^{close} = \left\{ (\theta_1, \theta_2) \in (\mathbb{R}^{\ell \times d})^2 \times (\mathbb{R}^{\ell \times d})^2 \mid \mathbf{x}^{\theta_1, \theta_2} \in \mathcal{X} \right\}, \tag{54}$$

where the state process $\mathbf{x}^{\theta_1, \theta_2} = (x_t^{\theta_1, \theta_2})_{t \geq 0}$ is controlled by the pair of closed-loop feedback control processes $(\mathbf{u}_1, \mathbf{u}_2) \in \mathcal{U}_{loc} \times \mathcal{U}_{loc}$ defined by

$$u_{i,t} = (-1)^i K_i (x_t^{\theta_1, \theta_2} - \bar{x}_t^{\theta_1, \theta_2}) + (-1)^i L_i \bar{x}_t^{\theta_1, \theta_2}. \tag{55}$$

When we plug in the above closed-loop feedback controls (55) into the state process dynamics (3), we obtain that, for every $t \geq 0$,

$$\begin{aligned} x_{t+1}^{\theta_1, \theta_2} = {} & (A - B_1 K_1 + B_2 K_2) \left(x_t^{\theta_1, \theta_2} - \bar{x}_t^{\theta_1, \theta_2} \right) \\ & + \left(\tilde{A} - \tilde{B}_1 L_1 + \tilde{B}_2 L_2 \right) \bar{x}_t^{\theta_1, \theta_2} + \epsilon_{t+1}^0 + \epsilon_{t+1}^1, \end{aligned}$$

where $\tilde{A} = A + \bar{A}$, $\tilde{B}_1 = B_1 + \bar{B}_1$, and $\tilde{B}_2 = B_2 + \bar{B}_2$. By Proposition 2, the process $\mathbf{x}^{\theta_1, \theta_2}$ is L^2-discounted globally integrable under the assumption:

$$\gamma \|A - B_1 K_1 + B_2 K_2\|^2 < 1, \qquad \text{and} \qquad \gamma \|\tilde{A} - \tilde{B}_1 L_1 + \tilde{B}_2 L_2\|^2 < 1. \tag{56}$$

Thus, it is reasonable to consider the following subset of Θ_{ad}^{close}:

$$\Theta = \left\{ (\theta_1, \theta_2) \in (\mathbb{R}^{\ell \times d})^2 \times (\mathbb{R}^{\ell \times d})^2 \mid (56) \text{ holds} \right\}. \tag{57}$$

One can check that the set Θ is not convex (see e.g. the Appendix of [38]). Moreover, the set Θ_{ad}^{close} does not have a simple expression in terms of the model parameters. Without any additional assumptions, the two players need to decide together the set of admissible policy profiles Θ_{ad}^{close} before playing against each other in a zero-sum game. However, in some situations, we can consider a subset of Θ_{ad}^{close} of the form $\Theta_1 \times \Theta_2$ where Θ_1, Θ_2 are two independent closed subsets in $\mathcal{U}_{ad}^{close}$, so that a player is able to choose freely and independently her admissible strategy without being affected by the L^2-integrability issue of the state process caused by the choice of strategy of her opponent.

Under Assumption 1, namely $\gamma\|A\|^2 < 1$ and $\gamma\|\tilde{A}\|^2 < 1$, there exists two pairs of real numbers $(\eta_1, \eta_2) \in \mathbb{R}^2$ and $(\tilde{\eta}_1, \tilde{\eta}_2) \in \mathbb{R}^2$ such that

$$\kappa := \gamma\|A\|^2 + \gamma\left(\eta_1^2\|B_1\|^2 + \eta_2^2\|B_2\|^2\right) < 1, \tilde{\kappa} := \gamma\|\tilde{A}\|^2 + \gamma\left(\tilde{\eta}_1^2\|\tilde{B}_1\|^2 + \tilde{\eta}_2^2\|\tilde{B}_2\|^2\right) < 1.$$

For $i = 1, 2$, let us denote $r_K^{(i)} = \eta_i\sqrt{\frac{1}{2}\left(\frac{1}{\kappa} - 1\right)}$, and $r_L^{(i)} = \tilde{\eta}_i\sqrt{\frac{1}{2}\left(\frac{1}{\tilde{\kappa}} - 1\right)}$.

The following result provides an example in which the two players are able to choose their admissible strategies independently of each other.

Lemma 20. *Assuming the closed-loop feedback policies $\theta_1 = (K_1, L_1) \in \mathbb{R}^{\ell\times d} \times \mathbb{R}^{\ell\times d}$ and $\theta_2 = (K_2, L_2) \in \mathbb{R}^{\ell\times d} \times \mathbb{R}^{\ell\times d}$ satisfy $\|K_i\| \le r_K^{(i)}$ and $\|L_i\| \le r_L^{(i)}$ for $i = 1, 2$, then $(\theta_1, \theta_2) \in \Theta$.*

The proof relies on Cauchy-Schwarz inequality. If the context is clear, we omit in the sequel the superscript (θ_1, θ_2) in state processes $(x_t^{\theta_1,\theta_2})_{t\ge0}$.

5.2. **Auxiliary processes.** We will use the following re-parametrization:

$$y_t = x_t - \bar{x}_t, \qquad z_t = \bar{x}_t, \qquad t \ge 0.$$

We denote $\mathbf{y} = (y_t)_{t\ge0}$ and $\mathbf{z} = (z_t)_{t\ge0}$ the two auxiliary state processes derived from $\mathbf{x}$. For the sake of clarity, we introduce some new notations on the control processes using the sample re-parametrization method:

$$\begin{aligned} u_{1,t}^{(y)} &:= u_{1,t} - \bar{u}_{1,t} = -K_1 y_t, & u_{2,t}^{(y)} &:= u_{2,t} - \bar{u}_{2,t} = K_2 y_t, \\ u_{1,t}^{(z)} &:= \bar{u}_{1,t} = -L_1 z_t, & u_{2,t}^{(z)} &:= \bar{u}_{2,t} = L_2 z_t. \end{aligned}$$

The processes $(y_t)_{t\ge0}$ and $(z_t)_{t\ge0}$ defined in this way follow the dynamics

$$y_{t+1} = Ay_t + B_1 u_{1,t}^{(y)} + B_2 u_{2,t}^{(y)} + \epsilon_{t+1}^1, \qquad y_0 \sim \epsilon_0^1, \tag{58}$$

$$z_{t+1} = \tilde{A}z_t + \tilde{B}_1 u_{1,t}^{(z)} + \tilde{B}_2 u_{2,t}^{(z)} + \epsilon_{t+1}^0, \qquad z_0 \sim \epsilon_0^0, \tag{59}$$

where $\epsilon_0^0, \epsilon_0^1$ are random variables with distributions μ_0^0 and μ_0^1 respectively. We then observe that for every $t, t' \ge 0$, the random variables y_t and z_t are respectively $\mathcal{F}^1-$measurable and $\mathcal{F}^0-$measurable, and they are independent.

The running cost at time t defined by (2) can also expressed as:

$$\begin{aligned} &c(x_t, \bar{x}_t, u_{1,t}, \bar{u}_{1,t}, u_{2,t}, \bar{u}_{2,t}) \\ &\quad= y_t^\top Q y_t + (u_{1,t}^{(y)})^\top R_1 u_{1,t}^{(y)} - (u_{2,t}^{(y)})^\top R_2 u_{2,t}^{(y)} + z_t^\top \tilde{Q} z_t \\ &\qquad\quad + (u_{1,t}^{(z)})^\top \tilde{R}_1 u_{1,t}^{(z)} - (u_{2,t}^{(z)})^\top \tilde{R}_2 u_{2,t}^{(z)} \\ &\quad= c_y(y_t, u_{1,t}^{(y)}, u_{2,t}^{(y)}) + c_z(z_t, u_{1,t}^{(z)}, u_{2,t}^{(z)}), \end{aligned}$$

where $\tilde{Q} = Q + \bar{Q}$, $\tilde{R}_1 = R_1 + \bar{R}_1$, $\tilde{R}_2 = R_2 + \bar{R}_2$, and $c_y : \mathbb{R}^d \times \mathbb{R}^\ell \times \mathbb{R}^\ell \to \mathbb{R}$ and $c_z : \mathbb{R}^d \times \mathbb{R}^\ell \times \mathbb{R}^\ell \to \mathbb{R}$ are the running cost functions associated to $(y_t)_{t\ge0}$ and $(z_t)_{t\ge0}$ defined by

$$\begin{aligned} c_y(y_t, u_{1,t}^{(y)}, u_{2,t}^{(y)}) &:= y_t^\top Q y_t + (u_{1,t}^{(y)})^\top R_1 u_{1,t}^{(y)} - (u_{2,t}^{(y)})^\top R_2 u_{2,t}^{(y)} \\ c_z(z_t, u_{1,t}^{(z)}, u_{2,t}^{(z)}) &:= z_t^\top \tilde{Q} z_t + (u_{1,t}^{(z)})^\top \tilde{R}_1 u_{1,t}^{(z)} - (u_{2,t}^{(z)})^\top \tilde{R}_2 u_{2,t}^{(z)}. \end{aligned}$$

We denote by $C(\theta_1, \theta_2) = J(\mathbf{u}_1, \mathbf{u}_2)$ the utility function associated to a closed-loop feedback policy profile $(\theta_1, \theta_2) \in \Theta$. As what has been presented in the proof

of Proposition 39, we introduce two auxiliary utility functions $C_y(K_1, K_2, \tilde{y})$ and $C_z(L_1, L_2, \tilde{z})$ defined as

$$C_y(K_1, K_2, \tilde{y}) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t c_y(y_t, u_{1,t}^{(y)}, u_{2,t}^{(y)}) \,\Big|\, y_0 = \tilde{y}\right], \tag{60}$$

$$C_z(L_1, L_2, \tilde{z}) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t c_z(z_t, u_{1,t}^{(z)}, u_{2,t}^{(z)}) \,\Big|\, z_0 = \tilde{z}\right], \tag{61}$$

in which the control processes are $(u_{1,t}^{(y)}, u_{1,t}^{(z)}) = (-K_1 y_t, -L_1 z_t)$ and $(u_{2,t}^{(y)}, u_{2,t}^{(z)}) = (K_2 y_t, L_2 z_t)$. Lemma 3 shows that $\mathbf{x}_t^{\theta_1,\theta_2}$ is L^2-integrable if and only if $\mathbf{y}$ and $\mathbf{z}$ are L^2-integrable. With the new notations, we let

$$C(\theta_1, \theta_2) = \mathbb{E}_{\tilde{y}}[C_y(K_1, K_2, \tilde{y})] + \mathbb{E}_{\tilde{z}}[C_z(L_1, L_2, \tilde{z})]. \tag{62}$$

Now we can define the closed-loop saddle point for the zero-sum game.

Definition 21. A closed-loop feedback policy profile $(\theta_1^*, \theta_2^*) \in \Theta_{ad}^{close}$ with $\theta_1^* = (K_1^*, L_1^*)$ and $\theta_2^* = (K_2^*, L_2^*)$ is said to be a closed-loop saddle point for the zero-sum game (CLSP for short) if and only if

- For every $\theta_1 = (K_1, L_1) \in \mathbb{R}^{\ell\times d} \times \mathbb{R}^{\ell\times d}$ such that $(\theta_1, \theta_2^*) \in \Theta_{ad}^{close}$,
$$C(\theta_1, \theta_2^*) \geq C(\theta_1^*, \theta_2^*),$$
- And for every $\theta_2 = (K_2, L_2) \in \mathbb{R}^{\ell\times d} \times \mathbb{R}^{\ell\times d}$ such that $(\theta_1^*, \theta_2) \in \Theta_{ad}^{close}$,
$$C(\theta_1^*, \theta_2) \leq C(\theta_1^*, \theta_2^*).$$

Remark 22. Note that the state processes associated with $C(\theta_1, \theta_2^*)$, $C(\theta_1^*, \theta_2^*)$, and $C(\theta_1^*, \theta_2)$ are all different.

Remark 23. We can see that the process $(y_t)_{t\geq 0}$ is completely controlled by (K_1, K_2) or by $(\mathbf{u}_1^{(y)}, \mathbf{u}_2^{(y)})$, and likewise for the process $(z_t)_{t\geq 0}$ by (L_1, L_2) or by $(\mathbf{u}_1^{(z)}, \mathbf{u}_2^{(z)})$. Moreover, the noise processes associated with $(y_t)_{t\geq 0}$ and $(z_t)_{t\geq 0}$ are independent. So when the two players are at CLSP (θ_1^*, θ_2^*), and one of them, say controller 1, perturbs her policy with a $\theta_1 = (K_1, L_1)$ different from $\theta_1^* = (K_1^*, L_1^*)$, we can look separately at $\mathbb{E}_{\tilde{y}}[C_y(K_1, K_2^*, \tilde{y})] - \mathbb{E}_{\tilde{y}}[C_y(K_1^*, K_2^*, \tilde{y})]$, and $\mathbb{E}_{\tilde{z}}[C_z(L_1, L_2^*, \tilde{z})] - \mathbb{E}_{\tilde{z}}[C_z(L_1^*, L_2^*, \tilde{z})]$.

We introduce here two sets related to the admissible policies with respect to the processes $(y_t)_{t\geq 0}$ and $(z_t)_{t\geq 0}$. Let us denote by

$$\Theta_y = \big\{(K_1, K_2) \in \mathbb{R}^{\ell\times d} \times \mathbb{R}^{\ell\times d} \mid \mathbf{y} \in \mathcal{X}\big\}, \tag{63}$$

$$\Theta_z = \big\{(L_1, L_2) \in \mathbb{R}^{\ell\times d} \times \mathbb{R}^{\ell\times d} \mid \mathbf{z} \in \mathcal{X}\big\}, \tag{64}$$

where $\mathbf{y}$ and $\mathbf{z}$ are two processes following the dynamics (58) and (59). The processes $\mathbf{y}$ and $\mathbf{z}$ can be constructed without any prior knowledge from $\mathbf{x}$, and they are completely determined by the choice of matrix pairs (K_1, K_2) and (L_1, L_2). From Lemma 3, the set Θ_y (and similarly Θ_z) can be understood as the collection of pair of matrices consisting of the first (or second) elements in policies θ_1 and θ_2.

Definition 24. A pair of matrices $(K_1^*, K_2^*) \in \mathbb{R}^{\ell\times d} \times \mathbb{R}^{\ell\times d}$ is said to be a closed-loop feedback saddle point in Θ_y ($CLSP-y$ for short) if for every $\tilde{y} \in \mathbb{R}^d$, for every $K_1, K_2 \in \mathbb{R}^{\ell\times d}$ such that $(K_1, K_2^*) \in \Theta_y$ and $(K_1^*, K_2) \in \Theta_y$, we have

$$C_y(K_1^*, K_2, \tilde{y}) \leq C_y(K_1^*, K_2^*, \tilde{y}) \leq C_y(K_1, K_2^*, \tilde{y}). \tag{65}$$

A pair of matrices $(L_1^*, L_2^*) \in \mathbb{R}^{\ell\times d}\times\mathbb{R}^{\ell\times d}$ is said to be a closed-loop feedback saddle point in Θ_z ($CLSP-z$ for short) if for every $\tilde{z}\in\mathbb{R}^d$, for every $L_1, L_2 \in \mathbb{R}^{\ell\times d}$ such that $(L_1, L_2^*)\in\Theta_z$ and $(L_1^*, L_2)\in\Theta_z$, we have

$$C_z(L_1^*, L_2, \tilde{z}) \le C_z(L_1^*, L_2^*, \tilde{z}) \le C_z(L_1, L_2^*, \tilde{z}). \tag{66}$$

5.3. **Notations and useful lemmas.** We will use the following notations:

$$\begin{cases} \mathcal{M}(P) = \gamma A^\top PA - P + Q \\ \mathcal{L}_1(P) = \gamma A^\top PB_1, \qquad \mathcal{L}_2(P) = \gamma A^\top PB_2, \qquad \mathcal{L}_{12} = \gamma B_1^\top PB_2 \\ \mathcal{N}_1(P) = \gamma B_1^\top PB_1 + R_1, \qquad \mathcal{N}_2(P) = \gamma B_2^\top PB_2 - R_2. \end{cases} \tag{67}$$

5.4. **Algebraic Riccati equations.** We present here a few lemmas and some notations that will be useful to understand the closed-loop saddle point in Θ_y ($CLSP-y$). Since the processes $\mathbf{y}$ and $\mathbf{z}$ follow similar linear dynamics but with different coefficients, we omit the proof for lemmas corresponding to $CLSP-z$. We use the notation $\langle u, v\rangle$ to represent the product $u^\top v$ for two vectors in $\mathbb{R}^d$. Using the dynamics (58) for $(y_t)_{t\ge0}$, we obtain the following result.

Lemma 25. *For every symmetric matrix $P\in\mathcal{S}^d$, we have*

$$\begin{aligned} &\gamma^{t+1}\mathbb{E}\left[\langle Py_{t+1}, y_{t+1}\rangle\right] \\ =&\gamma^t\mathbb{E}\left[\langle Py_t, y_t\rangle\right] + \gamma^{t+1}\mathbb{E}\left[(\epsilon_{t+1}^1)^\top P\epsilon_{t+1}^1\right] \\ &+\gamma^t\mathbb{E}\left[\begin{bmatrix} y_t \\ u_{1,t}^{(y)} \\ u_{2,t}^{(y)} \end{bmatrix}^\top \begin{bmatrix} \mathcal{M}(P)-Q & \mathcal{L}_1(P) & \mathcal{L}_2(P) \\ \mathcal{L}_1(P)^\top & \mathcal{N}_1(P)-R_1 & \mathcal{L}_{12}(P) \\ \mathcal{L}_2(P)^\top & \mathcal{L}_{12}(P)^\top & \mathcal{N}_2(P)+R_2 \end{bmatrix} \begin{bmatrix} y_t \\ u_{1,t}^{(y)} \\ u_{2,t}^{(y)} \end{bmatrix}\right] \end{aligned}$$

where we recall the notation (67).

Let us denote

$$C_y^*(P;\tilde{y}) := \tilde{y}^\top P\tilde{y} + \frac{\gamma}{1-\gamma}\mathbb{E}\left[(\epsilon_1^1)^\top P\epsilon_1^1\right].$$

Corollary 26. *For every $P\in\mathcal{S}^d$ and every $(K_1, K_2)\in\Theta_y$, we have*

$$\begin{aligned} &C_y(K_1, K_2, \tilde{y}) - C_y^*(P;\tilde{y}) \\ &= \mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t \begin{bmatrix} y_t \\ u_{1,t}^{(y)} \\ u_{2,t}^{(y)} \end{bmatrix}^\top \begin{bmatrix} \mathcal{M}(P) & \mathcal{L}_1(P) & \mathcal{L}_2(P) \\ \mathcal{L}_1(P)^\top & \mathcal{N}_1(P) & \mathcal{L}_{12}(P) \\ \mathcal{L}_1(P)^\top & \mathcal{L}_{12}(P)^\top & \mathcal{N}_2(P) \end{bmatrix} \begin{bmatrix} y_t \\ u_{1,t}^{(y)} \\ u_{2,t}^{(y)} \end{bmatrix} \middle| y_0 = \tilde{y}\right]. \end{aligned} \tag{68}$$

Remark 27. Notice that the difference between $C_y(K_1, K_2, \tilde{y})$ and $C_y^*(P;\tilde{y})$ depends on the cross product $\langle\mathcal{L}_{12}(P)^\top u_{1,t}^{(y)}, u_{2,t}^{(y)}\rangle$. When we perturb only one policy parameter, say K_1 for example, the change involved in the cost $C_y(K_1, K_2, \tilde{y})$ is not only caused by the state process $(y_t)_{t\ge0}$ but also by the interactions between the two feedback control processes, even if no term in definition (60) of $C_y(K_1, K_2, \tilde{y})$ is directly related to this cross interaction between strategies. The cross product $\mathcal{L}_{12}(P)$ in equation (68) makes Proposition 32 for the CLSP harder to prove than in a continuous-time result as discussed e.g. in [54].

Proof. The process $\mathbf{y} = (y_t)_{t\ge0}$ with dynamics (58) where $(K_1, K_2)\in\Theta_y$ satisfies

$$y_{t+1} = (A - B_1K_1 + B_2K_2)y_t + \epsilon_{t+1}^1, \qquad y_0 = \tilde{y}.$$

By definition of Θ_y, see (63), $\mathbf{y}$ is L^2-discounted globally integrable. By the definition of L^2-asymptotical stability, $\lim_{t\to 0}\mathbb{E}[\gamma^t\|y_t\|^2]=0$. Thus,

$$\lim_{t\to\infty}\left|\mathbb{E}\left[\gamma^t\langle Py_t,y_t\rangle|y_0=\tilde{y}\right]\right|\leq\lim_{t\to\infty}\|P\|\mathbb{E}\left[\gamma^t\|y_t\|^2|y_0=\tilde{y}\right]=0.$$

By applying recursively Lemma 25 from $t=T$ down to 0, and then letting T tends to infinity, we obtain equation (68). □

We introduce here another discrete ARE in $\mathcal{S}^d$ for the discrete-time process $(y_t)_t$ and the cost C_y:

$$0=\mathcal{M}(P)-\mathcal{L}(P)\mathcal{N}(P)^{-1}\mathcal{L}(P)^\top \tag{69}$$

with (using the notations introduced in (67))

$$\mathcal{L}(P)=[\mathcal{L}_1(P),\mathcal{L}_2(P)]\in\mathbb{R}^{d\times(\ell+\ell)} \tag{70}$$

and the 2×2 block matrix

$$\mathcal{N}(P)=\begin{bmatrix}\mathcal{N}_1(P) & \mathcal{L}_{12}(P)\\ \mathcal{L}_{12}(P)^\top & \mathcal{N}_2(P)\end{bmatrix}\in\mathbb{R}^{(\ell+\ell)\times(\ell+\ell)}. \tag{71}$$

To distinguish with other AREs introduced earlier, we may refer (69) as (ARE-y).

In the spirit of Nash equilibria, we discuss in the following a few results related to the situations when only one controller intends to change her strategy.

Let us denote by $\mathbf{y}^{2*}=(y_t^{2*})_{t\geq 0}$ the state process associated to a pair of strategies $(K_1,K_2^*)\in\Theta_y$ where K_2^* is a given matrix in $\mathbb{R}^{\ell\times d}$. Then $\mathbf{y}^{2*}$ follows the dynamics

$$y_{t+1}^{2*}=(A+B_2K_2^*)y_t^{2*}+B_1u_{1,t}^{(y)}+\epsilon_{t+1}^1=(A+B_2K_2^*-B_1K_1)y_t^{2*}+\epsilon_{t+1}^1,\ y_0^{2*}=\tilde{y}, \tag{72}$$

where $(u_{1,t}^{(y)})_{t\geq 0}$ is the control process adopted by player 1 (with parameter K_1).

Corollary 28. *There holds*

$$\begin{aligned}&C_y(K_1,K_2^*,\tilde{y})-C_y^*(P;\tilde{y})\\&=\mathbb{E}\left[\sum_{t=0}^\infty\gamma^t\begin{bmatrix}y_t^{2*}\\ u_{1,t}^{(y)}\end{bmatrix}^\top\begin{bmatrix}\mathcal{M}^{2*}(P) & \mathcal{L}_1^{2*}(P)\\ \mathcal{L}_1^{2*}(P)^\top & \mathcal{N}_1^{2*}(P)\end{bmatrix}\begin{bmatrix}y_t^{2*}\\ u_{1,t}^{(y)}\end{bmatrix}\middle|\, y_0=\tilde{y}\right]\end{aligned}$$

where

$$\begin{cases}\mathcal{M}^{2*}(P)=\mathcal{M}(P)+\mathcal{L}_2(P)K_2^*+(\mathcal{L}_2(P)K_2^*)^\top+(K_2^*)^\top\mathcal{N}_2(P)K_2^*\\ \mathcal{L}_1^{2*}(P)=\gamma(A+B_2K_2^*)^\top PB_1=\mathcal{L}_1(P)+(\mathcal{L}_{12}(P)K_2^*)^\top\\ \mathcal{N}_1^{2*}(P)=\gamma B_1^\top PB_1+R_1=\mathcal{N}_1(P).\end{cases}$$

The ARE associated to $\mathbf{y}^{2*}$ and the value function $C_y(\cdot,K_2^*,\tilde{y})$ is given by

$$0=\mathcal{M}^{2*}(P)-\mathcal{L}_1^{2*}(P)(\mathcal{N}_1^{2*}(P))^{-1}(\mathcal{L}_1^{2*}(P))^\top. \tag{73}$$

With a proper choice of K_2^*, we can connect equation (73) to the (ARE-y).

Lemma 29. *For any symmetric matrix $P\in\mathcal{S}^d$ such that $\mathcal{N}_1(P)$ and $S_2=\mathcal{N}_2(P)-\mathcal{L}_{12}(P)^\top\mathcal{N}_1(P)^{-1}\mathcal{L}_{12}(P)$ are invertible, let us consider that player 2 fixes her strategy with parameter*

$$K_2^*=\left(\mathcal{L}_{12}(P)^\top\mathcal{N}_2(P)^{-1}\mathcal{L}_{12}(P)-\mathcal{N}_2(P)\right)^{-1}\left(\mathcal{L}_2(P)^\top-\mathcal{L}_{12}(P)^\top\mathcal{N}_1(P)^{-1}\mathcal{L}_1(P)^\top\right). \tag{74}$$

Then P is a solution to (73) if and only if it is a solution to the (ARE-y) (69).

Proof. To alleviate the notations, we omit the matrix P in this proof. First, we notice that

$$K_2^* = -S_2^{-1}\Big(\mathcal{L}_2^\top - \mathcal{L}_{12}^\top \mathcal{N}_1^{-1}\mathcal{L}_1^\top\Big).$$

Then, the right hand side of ARE (73) becomes:

$$\begin{aligned}\mathcal{M}^{2*} - \mathcal{L}_1^{2*}(\mathcal{N}_1^{2*})^{-1}(\mathcal{L}_1^{2*})^\top &= \mathcal{M} - \mathcal{L}_1\mathcal{N}_1^{-1}\mathcal{L}_1^\top + \Big(\mathcal{L}_2 - \mathcal{L}_1\mathcal{N}_1^{-1}\mathcal{L}_{12}\Big)K_2^* \\ &\quad + (K_2^*)^\top\Big(\mathcal{L}_2^\top - \mathcal{L}_{12}^\top\mathcal{N}_1^{-1}\mathcal{L}_1^\top\Big) + (K_2^*)^\top\Big(\mathcal{N}_2 - \mathcal{L}_{12}^\top\mathcal{N}_1^{-1}\mathcal{L}_{12}\Big)K_2^*.\end{aligned}$$

Since $K_2^* = -S_2^{-1}\Big(\mathcal{L}_2^\top - \mathcal{L}_{12}^\top\mathcal{N}_1^{-1}\mathcal{L}_1^\top\Big)$ and $S_2 = \mathcal{N}_2 - \mathcal{L}_{12}^\top\mathcal{N}_1^{-1}\mathcal{L}_{12}$, we have

$$\begin{aligned}&\mathcal{M}^{2*} - \mathcal{L}_1^{2*}(\mathcal{N}_1^{2*})^{-1}(\mathcal{L}_1^{2*})^\top \\ &= \mathcal{M} - \mathcal{L}_1\mathcal{N}_1^{-1}\mathcal{L}_1^\top - \Big(\mathcal{L}_2 - \mathcal{L}_1\mathcal{N}_1^{-1}\mathcal{L}_{12}\Big)S_2^{-1}\Big(\mathcal{L}_2 - \mathcal{L}_1\mathcal{N}_1^{-1}\mathcal{L}_{12}\Big)^\top.\end{aligned}$$

Using the invertibility of $\mathcal{N}_1$ and S_2, we apply [48, Corollary 4.1] to $\mathcal{N}$ and obtain

$$\begin{aligned}&\mathcal{M} - \mathcal{L}\mathcal{N}^{-1}\mathcal{L}^\top \\ &= \mathcal{M} - [\mathcal{L}_1, \mathcal{L}_2]\begin{bmatrix} \mathcal{N}_1^{-1} + \mathcal{N}_1^{-1}\mathcal{L}_{12}S_2^{-1}\mathcal{L}_{12}^\top\mathcal{N}_1^{-1} & -\mathcal{N}_1^{-1}\mathcal{L}_{12}S_2^{-1} \\ -S_2^{-1}L_{12}^\top\mathcal{N}_1^{-1} & S_2^{-1}\end{bmatrix}\begin{bmatrix}\mathcal{L}_1^\top \\ \mathcal{L}_2^\top\end{bmatrix} \\ &= \mathcal{M} - \mathcal{L}_1\mathcal{N}_1^{-1}\mathcal{L}_1^\top - \Big(\mathcal{L}_2 - \mathcal{L}_1\mathcal{N}_1^{-1}\mathcal{L}_{12}\Big)S_2^{-1}\Big(\mathcal{L}_2 - \mathcal{L}_1\mathcal{N}_1^{-1}\mathcal{L}_{12}\Big)^\top.\end{aligned}$$

Hence, $\mathcal{M}^{2*} - \mathcal{L}_1^{2*}(\mathcal{N}_1^{2*})^{-1}(\mathcal{L}_1^{2*})^\top = \mathcal{M} - \mathcal{L}\mathcal{N}^{-1}\mathcal{L}^\top$. □

We state here briefly the counterparts of Corollary 28 and Lemma 29 for the situation when player 1 fixed her strategy to some predetermined matrix $K_1^* \in \mathbb{R}^{\ell\times d}$.

Corollary 30.

$$\begin{aligned}&C_y(K_1^*, K_2, \tilde{y}) - C_y^*(P;\tilde{y}) \\ &= \mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t\begin{bmatrix} y_t^{1*} \\ u_{2,t}^{(y)}\end{bmatrix}^\top\begin{bmatrix}\mathcal{M}^{1*}(P) & \mathcal{L}_2^{1*}(P) \\ \mathcal{L}_2^{1*}(P)^\top & \mathcal{N}_2^{1*}(P)\end{bmatrix}\begin{bmatrix} y_t^{1*} \\ u_{2,t}^{(y)}\end{bmatrix}\,\middle|\, y_0 = \tilde{y}\right],\end{aligned}$$

where

$$\begin{cases}\mathcal{M}^{1*}(P) = \mathcal{M}(P) - \mathcal{L}_1(P)K_1^* - (\mathcal{L}_1(P)K_1^*)^\top + (K_1^*)^\top\mathcal{N}_1(P)K_1^* \\ \mathcal{L}_2^{1*}(P) = \gamma(A - B_1K_1^*)^\top PB_2 = \mathcal{L}_2(P) - (\mathcal{L}_{12}(P)^\top K_1^*)^\top \\ \mathcal{N}_2^{1*}(P) = \gamma B_2^\top PB_2 - R_2 = \mathcal{N}_2(P),\end{cases}$$

and the state process $(y_t^{1*})_{t\geq 0}$ follows the dynamics

$$y_{t+1}^{1*} = (A - B_1K_1^*)y_t^{1*} + B_2u_{2,t}^{(y)} + \epsilon_{t+1}^1, \qquad y_0^{1*} = \tilde{y}. \tag{75}$$

Lemma 31. *If player 1 chooses her strategy with parameter*

$$K_1^* = -\Big(\mathcal{L}_{12}(P)\mathcal{N}_2(P)^{-1}\mathcal{L}_{12}(P)^\top - \mathcal{N}_1(P)\Big)^{-1}\Big(\mathcal{L}_1(P)^\top - \mathcal{L}_{12}(P)\mathcal{N}_2(P)^{-1}\mathcal{L}_2(P)^\top\Big),$$

where $P \in \mathcal{S}^d$ and the matrices $\mathcal{N}_2(P)$ and $S_1 = \mathcal{N}_1(P) - \mathcal{L}_{12}(P)\mathcal{N}_2(P)^{-1}\mathcal{L}_{12}(P)^\top$ are invertible, then we have

$$\mathcal{M}^{1*}(P) - \mathcal{L}_2^{1*}(P)(\mathcal{N}_2^{1*}(P))^{-1}(\mathcal{L}_2^{1*}(P))^\top = \mathcal{M}(P) - \mathcal{L}(P)\mathcal{N}(P)^{-1}\mathcal{L}(P)^\top.$$

5.5. **Sufficient condition.** We now phrase a sufficient condition of optimality. The necessary part will be discussed in Section 7, since it will serve as a basis for our numerical algorithms. For a symmetric matrix $P \in \mathcal{S}^d$, let us denote

$$K_i^* = -\Big(\mathcal{L}_{12}(P)\mathcal{N}_j(P)^{-1}\mathcal{L}_{12}(P)^\top - \mathcal{N}_i(P)\Big)^{-1}\Big(\mathcal{L}_i(P)^\top - \mathcal{L}_{12}(P)\mathcal{N}_j(P)^{-1}\mathcal{L}_j(P)^\top\Big) \tag{76}$$

for $i \neq j$, $i, j \in \{1, 2\}$, provided the inverse of matrices involved above exist.

Proposition 32. *Assume that we have the following two conditions:*

1. *The (ARE-y) (69) admits a symmetric solution $P \in \mathcal{S}^d$ satisfying*
$$\gamma B_1^\top P B_1 + R_1 \succ 0, \qquad \gamma B_2^\top P B_2 - R_2 \prec 0. \tag{77}$$
2. *The pair of matrices $(K_1^*, K_2^*) \in \Theta_y$.*

Then (K_1^*, K_2^*) is a $CLSP - y$ in Θ_y. Moreover, we have

$$C_y(K_1^*, K_2^*, \tilde{y}) = C_y^*(P; y) = \tilde{y}^\top P \tilde{y} + \frac{\gamma}{1-\gamma}\mathbb{E}\left[(\epsilon_1^1)^\top P \epsilon_1^1\right].$$

The control processes $(u_{1,t}^{(y),*})_{t\geq 0}$ and $(u_{1,t}^{(y),*})_{t\geq 0}$ corresponding to the $CLSP - y$ (K_1^*, K_2^*) are given by, for every $t \geq 0$,

$$u_{1,t}^{(y),*} = -K_1^* y_t^*, \qquad u_{2,t}^{(y),*} = K_2^* y_t^* \tag{78}$$

where the process $(y_t^*)_{t\geq 0}$ follows the dynamics

$$y_{t+1}^* = (A - B_1 K_1^* + B_2 K_2^*) y_t^* + \epsilon_{t+1}^1, \qquad y_0^* = \tilde{y}.$$

These two control processes satisfy the optimality condition: for every $t \geq 0$,

$$\mathcal{N}(P) u_t^{(y),*} + \mathcal{L}(P)^\top y_t^* = 0 \tag{79}$$

where $u_t^{(y),*} = [(u_{1,t}^{(y),*})^\top, (u_{2,t}^{(y),*})^\top]^\top \in \mathbb{R}^{2\ell}$, or equivalently

$$\mathcal{N}_1(P) u_{1,t}^{(y),*} + \mathcal{L}_{12}(P) u_{2,t}^{(y),*} = -\mathcal{L}_1(P)^\top y_t^*, \tag{80}$$

$$\mathcal{L}_{12}^\top(P) u_{1,t}^{(y),*} + \mathcal{N}_2(P) u_{2,t}^{(y),*} = -\mathcal{L}_2(P)^\top y_t^*. \tag{81}$$

Proof. From condition (77), $\mathcal{N}_1(P) \succ 0$ and $\mathcal{L}_{12}(P)^\top \mathcal{N}_1(P)^{-1}\mathcal{L}_{12}(P) - \mathcal{N}_2(P) \succ 0$, so the matrix K_2^* is well defined in $\mathbb{R}^{\ell\times d}$. Similarly, we have $\mathcal{N}_2(P) \prec 0$ and $\mathcal{L}_{12}(P)\mathcal{N}_2(P)^{-1}\mathcal{L}_{12}(P)^\top - \mathcal{N}_1(P) \prec 0$, so that K_1^* is well-defined too. Moreover, [48, Corollary 4.1] implies that the 2×2 block matrix $\mathcal{N}(P)$ defined by (71) is invertible. Applying Schur's lemma to the block matrix in (68), we get: for $t \geq 0$,

$$\begin{aligned}
&\begin{bmatrix} y_t \\ u_{1,t}^{(y)} \\ u_{2,t}^{(y)} \end{bmatrix}^\top \begin{bmatrix} \mathcal{M}(P) & \mathcal{L}_1(P) & \mathcal{L}_2(P) \\ \mathcal{L}_1(P)^\top & \mathcal{N}_1(P) & \mathcal{L}_{12}(P) \\ \mathcal{L}_1(P)^\top & \mathcal{L}_{12}(P)^\top & \mathcal{N}_2(P) \end{bmatrix} \begin{bmatrix} y_t \\ u_{1,t}^{(y)} \\ u_{2,t}^{(y)} \end{bmatrix} \\
=& y_t^\top \Big(\mathcal{M}(P) - \mathcal{L}(P)\mathcal{N}(P)^{-1}\mathcal{L}(P)^\top\Big) y_t \\
&+ \Big(u_t^{(y)} + \mathcal{N}(P)^{-1}\mathcal{L}(P)^\top y_t\Big)^\top \mathcal{N}(P)\Big(u_t^{(y)} + \mathcal{N}(P)^{-1}\mathcal{L}(P)^\top y_t\Big) \\
=& (i)_t + (ii)_t,
\end{aligned}$$

where $u_t^{(y)} = [(u_{1,t}^{(y)})^\top, (u_{2,t}^{(y)})^\top]^\top \in \mathbb{R}^{2\ell}$. Since P satisfies (ARE-y) (69), $(i)_t = 0$ for every $t \geq 0$. Thus, by Corollary 26, for every $(K_1, K_2) \in \Theta_y$ and $\tilde{y} \in \mathbb{R}^d$,

$$C_y(K_1, K_2, \tilde{y}) - C^*(P; \tilde{y})$$

$$= \mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t\Big(u_t^{(y)}+\mathcal{N}(P)^{-1}\mathcal{L}(P)^\top y_t\Big)^\top \mathcal{N}(P)\Big(u_t^{(y)}+\mathcal{N}(P)^{-1}\mathcal{L}(P)^\top y_t\Big)\,\Big|\, y_0=\tilde{y}\right].$$

If we choose $(K_1^*, K_2^*) \in \mathbb{R}^{\ell\times d}\times\mathbb{R}^{\ell\times d}$ satisfying equation (79), then we obtain $(ii)_t = 0$ for every $t \geq 0$. In this case, we have

$$C_y(K_1^*, K_2^*, \tilde{y}) = C_y^*(P;\tilde{y}) = \tilde{y}^\top P\tilde{y} + \frac{\gamma}{1-\gamma}\mathbb{E}\left[(\epsilon_1^1)^\top P\epsilon_1^1\right] < \infty.$$

Let us move on to obtain expressions for (K_1^*, K_2^*). Since the matrix $\mathcal{N}(P)$ is invertible, there exists a unique solution $u_t^{(y),*}$ to (79) for every $t \geq 0$. We plug-in the definition of $\mathcal{L}(P)$, $\mathcal{N}(P)$, and $u_t^{(y),*} = [(u_{1,t}^{(y),*})^\top, (u_{2,t}^{(y),*})^\top]^\top$, equation (79) is equivalent to the system of equations (80) and (81). So, by multiplying $\mathcal{L}_{12}(P)\mathcal{N}_2(P)^{-1}$ on both sides of (81), and subtract it to (80), we obtain

$$\Big(\mathcal{L}_{12}(P)\mathcal{N}_2(P)^{-1}\mathcal{L}_{12}(P)^\top - \mathcal{N}_1(P)\Big)u_{1,t}^{(y),*} = \Big(\mathcal{L}_1(P)^\top - \mathcal{L}_{12}(P)\mathcal{N}_2(P)^{-1}\mathcal{L}_2(P)^\top\Big)y_t^*.$$

From the assumptions, $\mathcal{L}_{12}(P)\mathcal{N}_2(P)^{-1}\mathcal{L}_{12}(P)^\top - \mathcal{N}_1(P) \prec 0$ is invertible. Hence, we obtain the optimal feedback control for player 1 by

$$u_{1,t}^{(y),*} = -K_1^* y_t^*$$

where K_1^* is given by (76). Similarly, we can derive that $u_{2,t}^{(y),*} = K_2^* y_t^*$ with K_2^* given by (76). Moreover, replacing $u_{1,t=0}^{(y),*}$ and $u_{2,t=0}^{(y),*}$ with their expressions in (78) back into (79), and by noticing that it holds true for every $\tilde{y} \in \mathbb{R}^d$, we have

$$\begin{cases} -\mathcal{N}_1(P)K_1^* + \mathcal{L}_{12}(P)K_2^* = -\mathcal{L}_1(P)^\top \\ -\mathcal{L}_{12}(P)^\top K_1^* + \mathcal{N}_2(P)K_2^* = -\mathcal{L}_2(P)^\top. \end{cases} \tag{82}$$

In the following, we will show that the pair (K_1^*, K_2^*) is a $CLSP-y$, which means that it satisfies condition (65). First, under the assumption in the statement, we know that $(K_1^*, K_2^*) \in \Theta_y$. Then, we look at the case when player 2 fixes her strategy to K_2^*, but player 1 adopts an alternative strategy K_1 satisfying $(K_1, K_2^*) \in \Theta_y$. The corresponding control at time t is then given by $u_{1,t}^{(y)} = -K_1 y_t^{2*}$, where $(y_t^{2*})_{t\geq 0}$ has dynamics (72). By Corollary 28 and Schur's lemma, we have

$$\begin{aligned} &C_y(K_1, K_2^*, \tilde{y}) - C_y^*(P;\tilde{y}) \\ =&\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t (y_t^{2*})^\top\Big(\mathcal{M}^{2*}(P) - \mathcal{L}_1^{2*}(P)(\mathcal{N}_1^{2*}(P))^{-1}\mathcal{L}_1^{2*}(P)^\top\Big)y_t^{2*}\,\Big|\, y_0^{2*}=\tilde{y}\right] \\ &+\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t\Big(u_{1,t}^{(y)} + (\mathcal{N}_1^{2*}(P))^{-1}\mathcal{L}_1^{2*}(P)^\top y_t^{2*}\Big)^\top \mathcal{N}_1^{2*}(P)\right. \\ &\qquad\left.\Big(u_{1,t}^{(y)} + (\mathcal{N}_1^{2*}(P))^{-1}\mathcal{L}_1^{2*}(P)^\top y_t^{2*}\Big)\,\Big|\, y_0^{2*}=\tilde{y}\right]. \end{aligned}$$

From Lemma 29, we know that a solution P to (ARE-y) (69) is also a solution to (73). Moreover, we have $-\mathcal{N}_1(P)K_1^* + \mathcal{L}_{12}(P)K_2^* = -\mathcal{L}_1(P)^\top$ which implies, by definition of $\mathcal{N}_1^{2*}(P)$ and $\mathcal{L}_1^{2*}(P)$, that

$$-K_1^* = -(\mathcal{N}_1(P))^{-1}\Big(\mathcal{L}_1(P)^\top + \mathcal{L}_{12}(P)K_2^*\Big) = -(\mathcal{N}_1^{2*}(P))^{-1}\mathcal{L}_1^{2*}(P)^\top.$$

Thus, together with $u_{1,t}^{(y)} = -K_1 y_t^{2*}$ for every $t \geq 0$, we have

$$C_y(K_1, K_2^*, \tilde{y}) - C^*(P;\tilde{y})$$

$$= \mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t(y_t^{2*})^\top(K_1^*-K_1)^\top\mathcal{N}_1(P)(K_1^*-K_1)y_t^{2*}\middle| y_0^{2*}=\tilde{y}\right].$$

Consequently, the condition $\mathcal{N}_1(P)=\gamma B_1^\top PB_1+R_1\succ 0$ implies

$$C_y(K_1,K_2^*,\tilde{y})-C_y(K_1^*,K_2^*,\tilde{y})\geq 0.$$

We can proceed similarly to prove that $C_y(K_1^*,K_2,\tilde{y})-C_y(K_1^*,K_2^*,\tilde{y})\leq 0$ using the fact that $K_2^*=-(\mathcal{N}_2^{1*}(P))^{-1}\mathcal{L}_2^{1*}(P)^\top$ and $(y_t^{1*})_{t\geq 0}$ satisfies the dynamics (75). □

We present here similar results corresponding to the $CLSP-z$. Let the matrices $(\tilde{N}_1,\tilde{N}_2,\tilde{L}_1,\tilde{L}_2,\tilde{L}_{12},\tilde{M},\tilde{N},\tilde{L})(\bar{P})$ be defined by using the same expressions in equations (67)(a), (b), (c), but by replacing (A,B_1,B_2,Q) to $(\tilde{A},\tilde{B}_1,\tilde{B}_2,\tilde{Q})$.

For a symmetric matrix $\bar{P}\in\mathcal{S}^d$, let us denote

$$L_i^*=-\Big(\tilde{\mathcal{L}}_{12}(\bar{P})\tilde{\mathcal{N}}_j(\bar{P})^{-1}\tilde{\mathcal{L}}_{12}(\bar{P})^\top-\tilde{\mathcal{N}}_i(\bar{P})\Big)^{-1}\Big(\tilde{\mathcal{L}}_i(\bar{P})^\top-\tilde{\mathcal{L}}_{12}(\bar{P})\tilde{\mathcal{N}}_j(\bar{P})^{-1}\tilde{\mathcal{L}}_j(\bar{P})^\top\Big) \tag{83}$$

for $i\neq j$, $i,j\in\{1,2\}$, provided the inverse of matrices appearing above exist.

Lemma 33. *Assume the following Algebraic Riccati equation (ARE-z):*

$$0=\tilde{\mathcal{M}}(\bar{P})-\tilde{\mathcal{L}}(\bar{P})\tilde{\mathcal{N}}(\bar{P})\tilde{\mathcal{L}}(\bar{P})^\top \tag{84}$$

admits a solution $\bar{P}\in\mathcal{S}^d$ which is such that

$$\gamma\tilde{B}_1^\top\bar{P}\tilde{B}_1+\tilde{R}_1\succ 0 \qquad \text{and} \qquad \gamma\tilde{B}_2^\top\bar{P}\tilde{B}_2-\tilde{R}_2\prec 0. \tag{85}$$

Assume in addition that the pair of matrices (L_1^*,L_2^*) given by (83) is in Θ_z. Then, (L_1^*,L_2^*) is a $CLSP-z$ in Θ_z.

Corollary 34. *If the two pairs $(K_1^*,K_2^*)\in\Theta_y$ and $(L_1^*,L_2^*)\in\Theta_z$ defined in Lemmas 32 and 33 are $CLSP-y$ and $CLSP-z$ respectively, then $(\theta_1^*,\theta_2^*)\in\Theta_{ad}^{close}$ defined by $\theta_1^*=(K_1^*,L_1^*)$, and $\theta_2^*=(K_1^*,L_2^*)$, is a closed-loop saddle point for the zero-sum game. The optimal value of the utility function is given by*

$$C(\theta_1^*,\theta_2^*)=\mathbb{E}\left[\tilde{y}^\top P\tilde{y}\right]+\mathbb{E}\left[\tilde{z}^\top\bar{P}\tilde{z}\right]+\frac{\gamma}{1-\gamma}\mathbb{E}\left[(\epsilon_1^1)^\top P\epsilon_1^1+(\epsilon_1^0)^\top\bar{P}\epsilon_1^0\right],$$

where P and $\bar{P}$ are solutions to the (ARE-y) (69) and (ARE-z) (84) satisfying conditions (77) and (85) respectively.

6. **Connection between closed-loop and open-loop saddle points.** In this section, we show that the open-loop and closed-loop equilibria are tightly related. To this end, we impose the following assumption on the model parameters.

Assumption 2. *We assume that $\ell=d$, and the matrices B_1,B_2,R_1,R_2 and $\tilde{B}_1,\tilde{B}_2,\tilde{R}_1,\tilde{R}_2$ are all invertible.*

For an invertible matrix $S\in\mathbb{R}^{d\times d}$, we denote $S^{-\top}=(S^\top)^{-1}=(S^{-1})^\top$.

If a solution $P^c\in\mathcal{S}^d$ to (ARE-y) is invertible, we have an alternative expression for (ARE-y).

Lemma 35. *Suppose Assumption 2 holds. Let $P^c\in\mathcal{S}^d$ be a solution to the ARE (69). If P^c and $(P^c)^{-1}+\gamma B_1R_1^{-1}B_1^\top-\gamma B_2R_2^{-1}B_2^\top$ are invertible, then*

$$\mathcal{M}(P^c)-\mathcal{L}(P^c)\mathcal{N}(P^c)^{-1}\mathcal{L}(P^c)^\top$$

$$= Q - P^c + A^\top \left(\frac{1}{\gamma}(P^c)^{-1} + B_1 R_1^{-1} B_1^\top - B_2 R_2^{-1} B_2^\top \right)^{-1} A. \tag{86}$$

Proof. First, under Assumption 2, we observe that

$$\begin{aligned} S_3 &:= \mathcal{L}_{12}(P^c)^\top - \mathcal{N}_2(P^c)\mathcal{L}_{12}(P^c)^{-1}\mathcal{N}_1(P^c) \\ &= \gamma B_2^\top P^c B_1 - (\gamma B_2^\top P^c B_2 - R_2)\left(\frac{1}{\gamma} B_2^{-1}(P^c)^{-1} B_1^{-\top}\right)(\gamma B_1^\top P^c B_1 + R_1) \\ &= R_2^\top B_2^{-1}\left(-B_2 R_2^{-1} B_2^\top + \frac{1}{\gamma}(P^c)^{-1} + B_1 R_1^{-1} B_1^\top\right) B_1^{-\top} R_1 \end{aligned} \tag{87}$$

and also

$$\begin{cases} \mathcal{N}_2(P^c)\mathcal{L}_{12}(P^c)^{-1} &= B_2^\top B_1^{-\top} - \frac{1}{\gamma} R_2 B_2^{-1}(P^c)^{-1} B_1^{-\top} \\ \mathcal{L}_{12}(P^c)^{-1}\mathcal{N}_1(P^c) &= \frac{1}{\gamma} B_2^{-1}(P^c)^{-1} B_1^{-T} R_1 + B_2^{-1} B_1. \end{cases} \tag{88}$$

Since $\mathcal{L}_{12}(P^c)$ and S_3 are invertible, by [48, Corollary 4.1] we obtain

$$\begin{aligned} &\frac{1}{\gamma^2}\mathcal{L}(P^c)\mathcal{N}(P^c)^{-1}\mathcal{L}(P^c)^\top \\ &= -A^\top P^c B_1 S_3^{-1}\mathcal{N}_2\mathcal{L}_{12}^{-1} B_1^\top P^c A + A^\top P^c B_1 S_3^{-1} B_2^\top P^c A \\ &\quad + A^\top P^c B_2 \left(\mathcal{L}_{12}^{-1} + \mathcal{L}_{12}^{-1}\mathcal{N}_1 S_3^{-1}\mathcal{N}_2\mathcal{L}_{12}^{-1}\right) B_1^\top P^c A - A^\top P^c B_2 \mathcal{L}_{12}^{-1}\mathcal{N}_1 S_3^{-1} B_2^\top P^c A \\ &= (i) + (ii) + (iii) + (iv). \end{aligned} \tag{89}$$

We then use equation (88) to simplify (i) and (iv):

$$(i) = -(ii) + \frac{1}{\gamma} A^\top P^c B_1 S_3^{-1} R_2 B_2^{-1} A, (iv) = -(ii) - \frac{1}{\gamma} A^\top B_1^{-T} R_1 S_3^{-1} B_2^\top P^c A.$$

Moreover we have

$$\begin{aligned} (iii) = \frac{1}{\gamma} A^\top P^c A + (ii) - \frac{1}{\gamma^2} A^\top B_1^{-\top} R_1 S_3^{-1} R_2 B_2^{-1} A + \frac{1}{\gamma} A^\top B_1^{-T} R_1 S_3^{-1} B_2^\top P^c A \\ - \frac{1}{\gamma} A^\top P^c B_1 S_3^{-1} R_2 B_2^{-1} A. \end{aligned}$$

Then, equation (89) becomes

$$\mathcal{L}(P^c)\mathcal{N}(P^c)^{-1}\mathcal{L}(P^c)^\top = \gamma A^\top P^c A - A^\top B_1^{-\top} R_1 S_3^{-1} R_2 B_2^{-1} A.$$

Together with equation (87), we conclude that

$$\begin{aligned} 0 &= \mathcal{M}(P^c) - \mathcal{L}(P^c)\mathcal{N}(P^c)^{-1}\mathcal{L}(P^c)^\top \\ &= Q - P^c + A^\top \left(\frac{1}{\gamma}(P^c)^{-1} + B_1 R_1^{-1} B_1^\top - B_2 R_2^{-1} B_2^\top \right)^{-1} A. \end{aligned}$$

□

Lemma 36. *Assume that Assumption 2 holds. Let $P^o \in \mathbb{R}^{d\times d}$ be a solution to the ARE (34) derived from the open-loop information structure, namely:*

$$P^o = \gamma \left(A^\top P^o + 2Q\right)\left(A + (B_1\Gamma_1 + B_2\Gamma_2)P^o\right) \tag{90}$$

where $\Gamma_1 = -\frac{1}{2}R_1^{-1}B_1^\top$ and $\Gamma_2 = \frac{1}{2}R_2^{-1}B_2^\top$. We consider the matrix given by

$$P^c = \frac{1}{2}A^\top P^o + Q. \tag{91}$$

If $A^\top P^o = (P^o)^\top A$, *and* P^c *and* $(P^c)^{-1}+\gamma\left(B_1R_1^{-1}B_1^\top - B_2R_2^{-1}B_2^\top\right)$ *are invertible, then* P^c *is a solution to (ARE-y)* (69).

Proof. This is a direct consequence of Lemma 35. By plugging the expressions of Γ_1 and Γ_2 into equation (90) and replacing $\frac{1}{2}A^\top P^o + Q$ by P^c, under the invertibility condition on P^c and $(P^c)^{-1} + \gamma\left(B_1R_1^{-1}B_1^\top - B_2R_2^{-1}B_2^\top\right)$, we get

$$\frac{1}{2}P^o = \left(\frac{1}{\gamma}(P^c)^{-1} + B_1R_1^{-1}B_1^\top - B_2^\top R_2^{-1}B_2^\top\right)^{-1} A.$$

Multiplying both sides by $A^\top$ and rearranging the terms, we obtain (86). □

Remark 37. In addition to Assumption 2, if A is invertible, then from a positive definite solution P^c to (ARE-y) (69), we can define $P^o = 2A^{-\top}(P^c - Q) \in \mathbb{R}^{d\times d}$. By inverting the steps used in Lemma 36, we can show that P^o solves (34).

The following corollary shows that both the pair of control processes associated to a closed-loop saddle point and the pair of processes for an OLSP will lead to the same state process, hence the same value function for the zero-sum game.

Corollary 38. *We assume that Assumption 2 holds and* A, $\tilde{A}$ *are invertible. Suppose that there exists unique invertible solutions* P^o *(resp.* $\bar{P}^o$*) and* P^c *(resp.* $\bar{P}^c$*) to the corresponding ARE in the open-loop information structure* (34) *(resp.*(35)*) and in the closed-loop information structure* (69) *(resp.* (84)*). Then, we have :*

(i) The following holds, where $((K_1^*, L_1^*), (K_2^*, L_2^*))$ *are given in* (76) *and* (83)*:*

$$\begin{cases} -B_1K_1^* + B_2K_2^* = B_1\Gamma_1P^o + B_2\Gamma_2P^o \\ -\tilde{B}_1L_1^* + \tilde{B}_2L_2^* = \tilde{B}_1\Lambda_1\bar{P}^o + \tilde{B}_2\Lambda_2\bar{P}^o. \end{cases} \tag{92}$$

(ii) For every time $t \geq 0$*, the state variable* $x_t^{\theta_1^*,\theta_2^*}$ *corresponding to a pair of closed-loop feedback control* $(\mathbf{u}_1^{c,*}, \mathbf{u}_2^{c,*})$ *with policies* $(\theta_1^*, \theta_2^*) = ((K_1^*, L_1^*), (K_2^*, L_2^*))$ (55) *has the same distribution as the state variable* $x_t^{\mathbf{u}_1^{o,*},\mathbf{u}_2^{o,*}}$ *controlled by an OLSP* $(\mathbf{u}_1^{o,*}, \mathbf{u}_2^{o,*})$ *with parameters* $(P^o, \bar{P}^o)$ (42).

Proof. According to Lemma 36, the unique solutions P^o (resp. $\bar{P}^o$) and P^c (resp. $\bar{P}^c$) to the corresponding Algebraic Riccai equation satisfy:

$$P^c = \frac{1}{2}A^\top P^o + Q, \qquad \text{and} \qquad \bar{P}^c = \frac{1}{2}(A+\bar{A})^\top\bar{P}^o + (Q+\bar{Q}).$$

(i) It is enough to show the connection between (K_1^*, K_2^*) to the pair of matrices $(-\Gamma_1P^o, -\Gamma_2P^o)$, and the situation for (L_1^*, L_2^*) can be proved with similar arguments. Let us denote by $\check{B} = [B_1, B_2] \in \mathbb{R}^{d\times 2\ell}$ and $\check{R} = \begin{bmatrix} R_1 & 0 \\ 0 & -R_2 \end{bmatrix}$. Then, by equation (82) and Lemma 35, since A is invertible, we have:

$$\begin{aligned} -B_1K_1^* + B_2K_2^* &= -\check{B}(\gamma\check{B}^\top P^c\check{B} + \check{R})^{-1}\left(\gamma\check{B}^\top P^cA\right) \\ &= -\left(\check{B}\check{R}^{-1}\check{B}^\top\right)\left(A^{-\top}(P^c - Q)\right). \end{aligned}$$

Together with $P^o = 2A^{-\top}(P^c - Q)$ and the definition of Γ_1, Γ_2, we obtain

$$-B_1K_1^* + B_2K_2^* = -\frac{1}{2}\check{B}\check{R}^{-1}\check{B}^\top P^o = B_1\Gamma_1P^o + B_2\Gamma_2P^o.$$

(ii) By comparing the state dynamics of $\left(x_t^{\mathbf{u}_1^{o,*},\mathbf{u}_2^{o,*}} - \bar{x}_t^{\mathbf{u}_1^{o,*},\mathbf{u}_2^{o,*}}\right)_{t\geq 0}$ in the open-loop case and that of $(y_t^{\theta_1^*,\theta_2^*})_{t\geq 0}$ (58) in the closed-loop case, we have $y_t^{\theta_1^*,\theta_2^*} \overset{d}{=}$

$x_t^{\mathbf{u}_1^{o,*},\mathbf{u}_2^{o,*}} - \bar{x}_t^{\mathbf{u}_1^{o,*},\mathbf{u}_2^{o,*}}$ in the sense of distribution. Similar arguments show $z_t^{\theta_1^*,\theta_2^*} \overset{d}{=} \bar{x}_t^{\theta_1^*,\theta_2^*}$. Thus, the conclusion holds. □

7. **Algorithms.** In this section, we propose policy-gradient based algorithms to find the Nash equilibrium of the zero-sum mean-field type game. We start with a convenient expression for the gradient of the utility function, which leads to a necessary condition of optimality (counterpart to the sufficient condition studied in § 5.5). Then, after introducing model-based methods, we explain how to extend them to sample-based algorithms in which the gradient is estimated using a simulator providing stochastic realizations of the utility. The results of this section have initially been presented in [25].

7.1. **Gradient expression.** We henceforth focus on the following problem based on closed-loop controls, introduced in Section 5. Each player $i = 1, 2$ chooses parameter $\theta_i^* = (K_i^*, L_i^*)$ such that $\theta = (\theta_1^*, \theta_2^*)$ is a CLSP (see Defintion 21 and (57) for the definition of the set Θ).

For simplicity, we introduce the following notation $x^{\mathbf{u}_1^{\theta_1},\mathbf{u}_2^{\theta_2}} = x^{\theta_1,\theta_2}$, and since we focus on linear controls, using the notation C introduced in (62), we have

$$C(\theta_1, \theta_2) = J(\mathbf{u}_1^{\theta_1}, \mathbf{u}_2^{\theta_2}).$$

Moreover, let $y_t^{K_1,K_2} = x_t^{\theta_1,\theta_2} - \bar{x}_t^{\theta_1,\theta_2}$ and $z_t^{L_1,L_2} = \bar{x}_t^{\theta_1,\theta_2}$, which is justified by the fact that the dynamics of $\mathbf{y}$ and $\mathbf{z}$ depend respectively only on (K_1, K_2) and (L_1, L_2).

Let $P^y_{K_1,K_2}$ and $P^z_{L_1,L_2}$ be a solutions to the linear equations

$$\begin{aligned} P^y_{K_1,K_2} &= Q + K_1^\top R_1 K_1 - K_2^\top R_2 K_2 \\ &\quad + \gamma(A - B_1K_1 + B_2K_2)^\top P^y_{K_1,K_2}(A - B_1K_1 + B_2K_2), \end{aligned} \tag{93}$$

$$\begin{aligned} P^z_{L_1,L_2} &= \tilde{Q} + L_1^\top \tilde{R}_1 L_1 - L_2^\top \tilde{R}_2 L_2 \\ &\quad + \gamma(\tilde{A} - \tilde{B}_1L_1 + \tilde{B}_2L_2)^\top P^z_{L_1,L_2}(\tilde{A} - \tilde{B}_1L_1 + \tilde{B}_2L_2). \end{aligned} \tag{94}$$

We now provide an explicit expression for the gradient of the utility function with respect to the control parameters in terms of the solution to the equations (93) and (94). Let us denote

$$\begin{bmatrix} E^{y,1}_{K_1,K_2} \\ E^{y,2}_{K_1,K_2} \end{bmatrix} = -\gamma \begin{bmatrix} B_1^\top P^y_{K_1,K_2} A \\ -B_2^\top P^y_{K_1,K_2} A \end{bmatrix} + \mathbf{R} \begin{bmatrix} K_1 \\ K_2 \end{bmatrix},$$

$$\begin{bmatrix} E^{z,1}_{L_1,L_2} \\ E^{z,2}_{L_1,L_2} \end{bmatrix} = -\gamma \begin{bmatrix} \tilde{B}_1^\top P^z_{L_1,L_2} \tilde{A} \\ -\tilde{B}_2^\top P^z_{L_1,L_2} \tilde{A} \end{bmatrix} + \tilde{\mathbf{R}} \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$$

with

$$\mathbf{R} = \begin{bmatrix} R_1 + \gamma B_1^\top P^y_{K_1,K_2} B_1 & -\gamma B_1^\top P^y_{K_1,K_2} B_2 \\ -\gamma B_2^\top P^y_{K_1,K_2} B_1 & -R_2 + \gamma B_2^\top P^y_{K_1,K_2} B_2 \end{bmatrix},$$

$$\tilde{\mathbf{R}} = \begin{bmatrix} \tilde{R}_1 + \gamma \tilde{B}_1^\top P^z_{L_1,L_2} \tilde{B}_1 & -\gamma \tilde{B}_1^\top P^z_{L_1,L_2} \tilde{B}_2 \\ -\gamma \tilde{B}_2^\top P^z_{L_1,L_2} \tilde{B}_1 & -\tilde{R}_2 + \gamma \tilde{B}_2^\top P^z_{L_1,L_2} \tilde{B}_2 \end{bmatrix}$$

where

$$\Sigma^y_{K_1,K_2} = \mathbb{E}\Big[\sum_{t\geq 0} \gamma^t y_t^{K_1,K_2} (y_t^{K_1,K_2})^\top\Big], \qquad \Sigma^z_{L_1,L_2} = \mathbb{E}\Big[\sum_{t\geq 0} \gamma^t z_t^{L_1,L_2} (z_t^{L_1,L_2})^\top\Big].$$

The following result has been proved in [25, Proposition 1].

Proposition 39 (Policy gradient expression)**.** *For any* $\theta = (\theta_1, \theta_2) \in \Theta$,

$$\nabla_{K_j} C(\theta_1, \theta_2) = 2E^{y,j}_{K_1,K_2} \Sigma^y_{K_1,K_2}, \quad \nabla_{L_j} C(\theta_1, \theta_2) \quad = 2E^{z,j}_{L_1,L_2} \Sigma^z_{L_1,L_2}, \quad j = 1, 2. \tag{95}$$

7.2. **Model-based policy optimization.** Let us assume that the model is known and both players can see the actions of one another at the end of each time step. To explain the intuition behind the iterative methods, we first express the optimal control of a player when the other player has a fixed control. For some given $\theta_2 = (K_2, L_2)$, the inner minimization problem for player 1 becomes an LQR problem with instantaneous utility at time t:

$$\begin{aligned}(x_t - \bar{x}_t)^\top \mathbf{Q}_{K_2} (x_t - \bar{x}_t) + \bar{x}^\top \tilde{\mathbf{Q}}_{K_2} \bar{x} \\ + (u_{1,t} - \bar{u}_{1,t})^\top R_1 (u_{1,t} - \bar{u}_{1,t}) + \bar{u}_{1,t}^\top (R_1 + \bar{R}_1) \bar{u}_{1,t},\end{aligned}$$

when player 1 uses control u_1, where $\mathbf{Q}_{K_2} = Q - K_2 R_2 K_2$ and $\tilde{\mathbf{Q}}_{L_2} = \tilde{Q} - L_2 \tilde{R}_2 L_2$, and state dynamics given by:

$$x_{t+1} = \mathbf{A}_{K_2} x_t + \bar{\mathbf{A}}_{K_2,L_2} \bar{x}_t + B_1 u_{1,t} + \bar{B}_1 \bar{u}_{1,t} + \epsilon^0_{t+1} + \epsilon^1_{t+1},$$

where $\mathbf{A}_{K_2} = A + B_2 K_2$ and $\bar{\mathbf{A}}_{K_2,L_2} = \bar{A} + \bar{B}_2 L_2 + B_2(L_2 - K_2)$. Inspired by the results in [38], we propose to find the stationary point $\theta_1^*(\theta_2) = (K_1^*(K_2), L_1^*(L_2))$ of the inner problem. By setting $\nabla_{\theta_1} C(\theta_1, \theta_2) = 0$ and by Proposition 39, this yields

$$K_1^*(K_2) = \gamma (R_1 + \gamma B_1^\top P^y_{K_2} B_1)^{-1} B_1^\top P^y_{K_2} \left[A + B_2 K_2\right], \tag{96}$$

where $P^y_{K_2} = P^y_{K_1^*(K_2),K_2}$ solves

$$P^y_{K_2} = \tilde{Q}_{K_2} + \gamma \tilde{A}^\top_{K_2} P^y_{K_2} \tilde{A}_{K_2} - \gamma^2 \tilde{A}^\top_{K_2} P^y_{K_2} B_1 (R_1 + \gamma B_1^\top P^y_{K_2} B_1)^{-1} B_1^\top P^y_{K_2} \tilde{A}_{K_2},$$

where $\tilde{Q}_{K_2} = Q - K_2^\top R_2 K_2$ and $\tilde{A}_{K_2} = A + B_2 K_2$. This equation is obtained by considering the equation (93) for $P^y_{K_1,K_2}$ and replacing K_1 by the above expression (96) for $K_1^*(K_2)$. One can similarly introduce $K_2^*(K_1)$, which is the optimal K_2 for a given K_1, and likewise for $L_1^*(L_2), L_2^*(L_1)$.

Based on this idea and inspired by the works of Fazel et al. [38] and Zhang et al. [57], we propose two iterative algorithms relying on policy-gradient methods, namely alternating-gradient and gradient-descent-ascent, to find the optimal values of θ_1 and θ_2. Starting from an initial guess of the control parameters, the players update either alternatively or simultaneously their parameters by following the gradients of the utility function. In the *alternating-gradient* (AG) method, the players take turn in updating their parameters. Between two updates of θ_2, θ_1 is updated $\mathcal{N}_1^{max}$ times. In the *gradient-descent-ascent* (GDA) method, all the control parameters are updated synchronously at each iteration. For description of the algorithms and more details, see e.g. [51] and [30, 32, 43, 49] respectively (see also [57]).

At each step of these methods, the gradients can be computed directly using the formulas provided in Proposition 39. In order to have a benchmark, one can compute the equilibrium (θ_1^*, θ_2^*) by solving the Riccati equations (69)–(84). Alternatively, the Nash equilibrium can be computed by finding K_2 such that $\nabla_{K_2} C_y(K_1^*(k_2), K_2)\big|_{k_2 = K_2} = 0$. The left-hand side has an explicit expression obtained by combining (95) and (96).

7.3. **Sample-based policy optimization.** The aforementioned methods use explicit expressions for the gradients, which rely on the knowledge of the model (the coefficients of the dynamics and the utility function). However, in many situations these coefficients are not known. Instead, let us assume that we have access to the following (stochastic) simulator, called *MKV simulator* and denoted by $\mathcal{S}^{\mathcal{T}}_{MKV}$: given a control parameter $\theta = (\theta_1, \theta_2) = (K_1, L_1, K_2, L_2)$, $\mathcal{S}^{\mathcal{T}}_{MKV}(\theta)$ returns a sample of the mean-field utility (i.e., the quantity inside the expectation in equation (4)) for the MKV dynamics (3) using the control θ and truncated at time horizon $\mathcal{T}$, which is similar to the one introduced in [28]. In other words, it returns a realization of the social utility $\sum_{t=0}^{\mathcal{T}-1} \gamma^t c_t$, where c_t is the instantaneous mean-field utility at time t, see (5). We can estimate the gradient of the utility with respect to the control parameters of each player. The estimation algorithm uses the simulator to obtain realizations of the (truncated) utility when using perturbed versions of the controls with $O(M)$ perturbations. See [28, 38] for more details

7.4. **Numerical results.** We now provide numerical results both for model-based and sample-based versions of the two methods presented in the previous section.

Setting. The specification of the model used in the simulations is given in Table 1. This setting has been chosen to illustrate the convergence when the equilibrium controls are not symmetric, i.e. $\theta_1 \neq \theta_2$. To be able to visualize the convergence of the controls, we focus on a one-dimensional example, that is, $d = \ell = 1$.

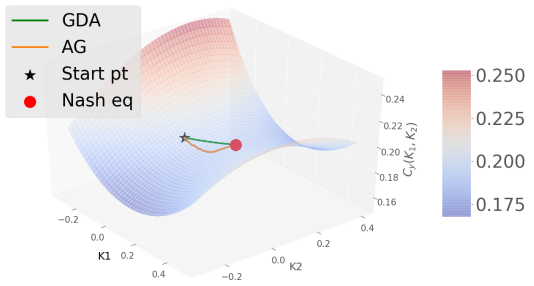
Results with exact gradients. The parameters used are given in Table 1, inspired by the values used for a single controller in [28] and numerical experiments.

Fig. 1 shows the trajectory of $(K_1, K_2) \mapsto C_y(K_1, K_2)$ and $(L_1, L_2) \mapsto C_z(L_1, L_2)$ generated by the iterations of AG and DGA methods. Iterations are counted in the following way: in AG at iteration k, $(\theta_1^k, \theta_2^k) = (\theta_1^{k \bmod \mathcal{N}_1^{max}, \lceil k/\mathcal{N}_1^{max} \rceil}, \theta_2^{\lceil k/\mathcal{N}_1^{max} \rceil - 1})$, while in DGA one step of for-loop corresponds to one iteration. The utility at the starting point and at the Nash equilibrium are respectively given by a black star and a red dot. In the AG, since θ_1 is updated $\mathcal{N}_1^{max}$ times between two updates of θ_2, the trajectory moves faster in the θ_1-direction until it reaches an approximate best response against θ_2, after which the trajectory moves towards the Nash equilibrium. This is also confirmed by the parameters convergence in Fig. 2(a). The relative error on the utility is shown in Fig. 2(b). We observe that the convergence is slower with AG because player 2 updates her control only every $\mathcal{N}_1^{max}$ iterations.

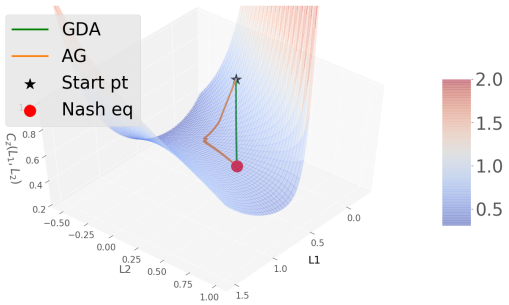
Sample-based results. The parameters used are given in Table 1 and were chosen based on the values in [28] as well as numerical experiments. The figures are obtained by averaging the results over 5 experiments, each based on a different realization of the randomness in the initial points, in the dynamics and in the gradient estimation.

Fig. 3 shows the trajectory of $(K_1, K_2) \mapsto C_y(K_1, K_2)$ and $(L_1, L_2) \mapsto C_z(L_1, L_2)$ generated by the iterations of AG and DGA methods. The convergence of the parameters $\theta = (K_1, L_1, K_2, L_2)$ is shown in Fig. 4(a). The evolution of the relative error on the utility is shown in Fig. 4(b).

8. **Conclusion.** In this paper, we have studied zero-sum mean-field type games with linear quadratic model under infinite-horizon discounted utility function. We have identified the closed-form expression of the Nash equilibrium controls as linear combinations of the state and its mean. Moreover, we have proposed two policy optimization methods to learn the equilibrium. Numerical results have shown the



(A)



(B)

FIGURE 1. Model-based policy optimization: Convergence of each part of the utility. (a) C_y as a function of (K_1, K_2). (b) C_z as a function of (L_1, L_2).

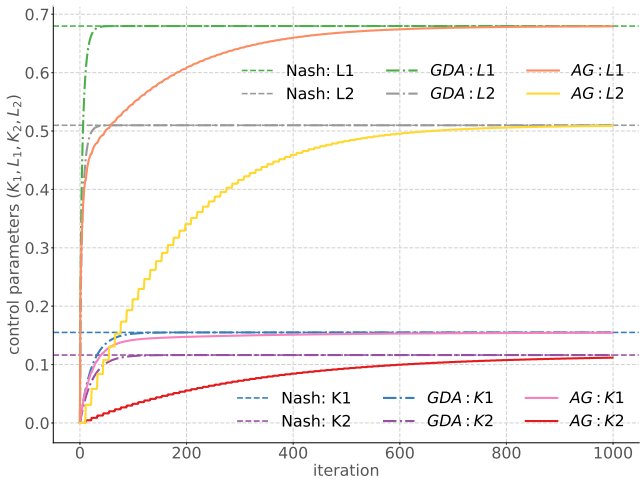
TABLE 1. Simulation parameters

Model parameters								
A	$\overline{A}$	$B_1 = \overline{B}_1$	$B_2 = \overline{B}_2$	Q	$\overline{Q}$	$R_1 = \overline{R}_1$	$R_2 = \overline{R}_2$	γ
0.4	0.4	0.4	0.3	0.4	0.4	0.4	0.4	0.9

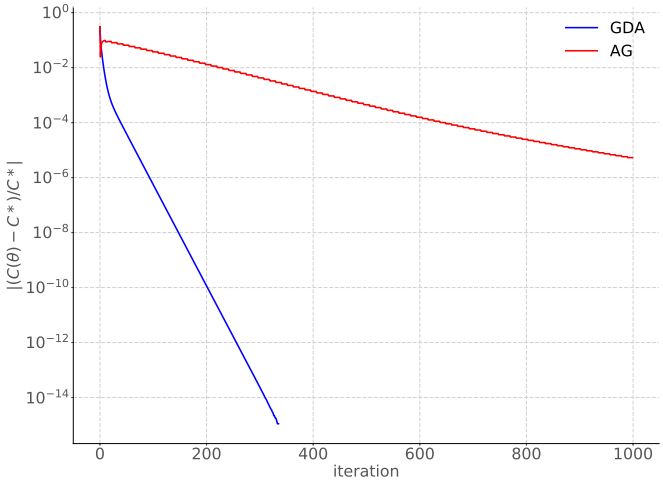
Initial distribution and noise processes			
ϵ_0^0	ϵ_0^1	ϵ_t^0	ϵ_t^1
$\mathcal{U}([-1,1])$	$\mathcal{U}([-1,1])$	$\mathcal{N}(0,0.01)$	$\mathcal{N}(0,0.01)$

AG and DGA methods parameters								
$\mathcal{N}_1^{max}$	$\mathcal{N}_2^{max}$	T	η_1	η_2	K_1^0	L_1^0	K_2^0	L_2^0
10	200	2000	0.1	0.1	0.0	0.0	0.0	0.0

Gradient estimation algorithm parameters		
$\mathcal{T}$	M	τ
50	10000	0.1



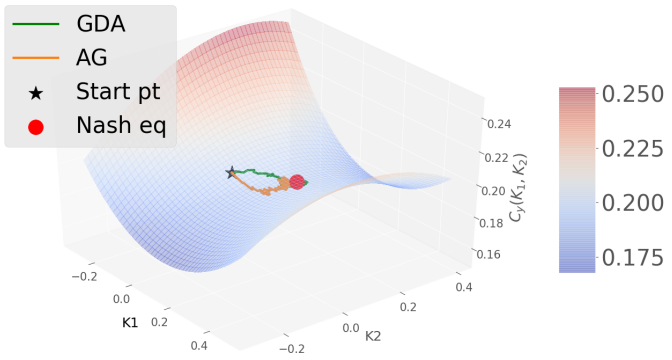
(A)



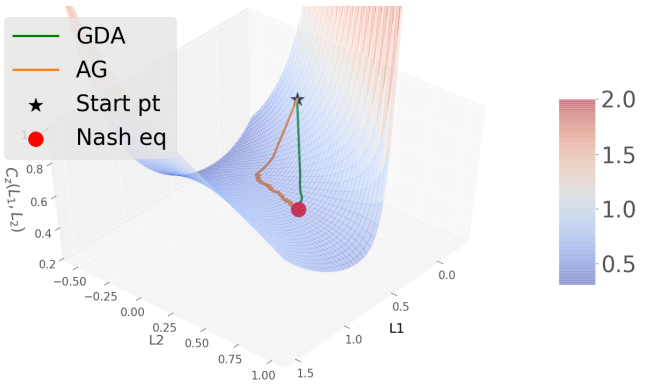
(B)

FIGURE 2. Model-based policy optimization: Convergence of the control parameters in (a) and of the relative error on the utility in (b).

convergence of the two methods in both model-based and sample-based settings. The question of convergence of the algorithms proposed here as well as model-free methods for non-LQ or general-sum MFTG will be studied in future works.

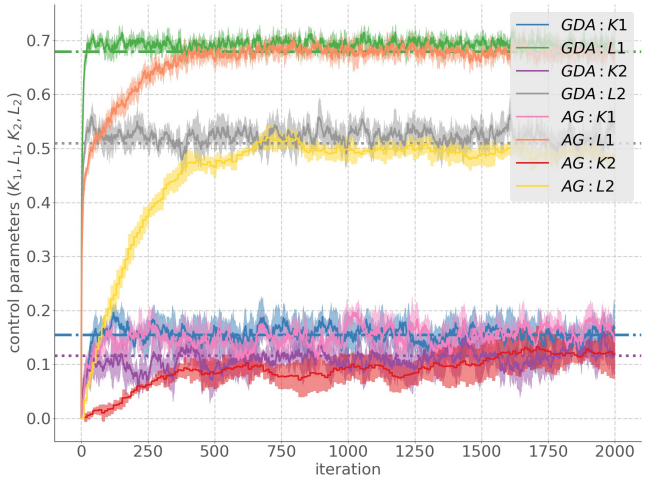


(A)

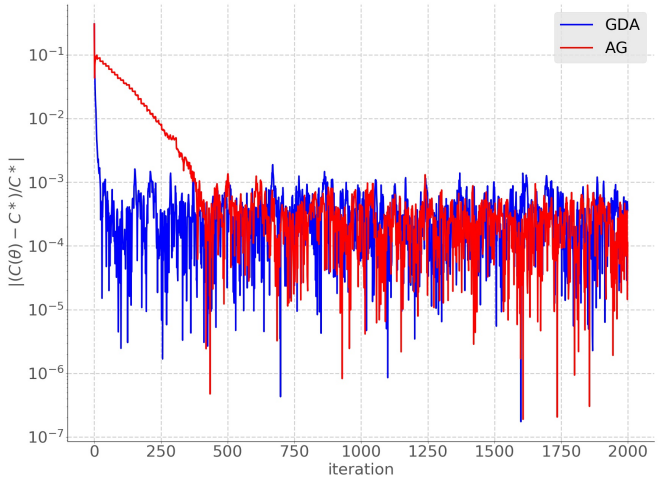


(B)

FIGURE 3. Sample-based policy optimization: Convergence of each part of the utility. (a) C_y as a function of (K_1, K_2). (b) C_z as a function of (L_1, L_2).



(A)



(B)

FIGURE 4. Sample-based policy optimization: Convergence of the control parameters in (a) and of the relative error on the utility in (b).

REFERENCES

[1] Y. Achdou, F. Camilli and I. Capuzzo-Dolcetta, Mean field games: Numerical methods for the planning problem, *SIAM J. Control Optim.*, **50** (2012), 77–109.

[2] Y. Achdou and I. Capuzzo-Dolcetta, Mean field games: Numerical methods, *SIAM J. Numer. Anal.*, **48** (2010), 1136–1162.

[3] Y. Achdou and J.-M. Lasry, Mean field games for modeling crowd motion, in *Contributions to Partial Differential Equations and Applications*, Comput. Methods Appl. Sci., 47, Springer, Cham, 2019, 17–42.

[4] Y. Achdou and M. Laurière, Mean field games and applications: Numerical aspects, in *Mean Field Games*, Lecture Notes in Math., 2281, Fond. CIME/CIME Found. Subser., Springer, Cham, 2020, 249–307.

[5] Y. Achdou and M. Laurière, Mean field type control with congestion (II): An augmented Lagrangian method, *Appl. Math. Optim.*, **74** (2016), 535–578.

[6] Y. Achdou and M. Laurière, On the system of partial differential equations arising in mean field type control, *Discrete Contin. Dyn. Syst.*, **35** (2015), 3879–3900.

[7] A. Al-Tamimi, F. L. Lewis and M. Abu-Khalaf, Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control, *Automatica J. IFAC*, **43** (2007), 473–481.

[8] C. Alasseur, I. Ben Tahar and A. Matoussi, An extended mean field game for storage in smart grids, *J. Optim. Theory Appl.*, **184** (2020), 644–670.

[9] B. Anahtarci, C. D. Karıksız and N. Saldi, Value iteration algorithm for mean-field games, *Systems Control Lett.*, **143** (2020), 10pp.

[10] J. Barreiro-Gomez, T. E. Duncan and H. Tembine, Discrete-time linear-quadratic mean-field-type repeated games: Perfect, incomplete, and imperfect information, *Automatica J. IFAC*, **112** (2020), 16pp.

[11] T. Başar and P. Bernhard, *H^∞ Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, Birkhäuser, Boston, MA, 2008.

[12] D. Bauso, *Game Theory with Engineering Applications*, Advances in Design and Control, 30, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2016.

[13] D. Bauso, H. Tembine and T. Başar, Robust mean field games with application to production of an exhaustible resource, *IFAC Proceedings Volumes*, **45** (2012), 454–459.

[14] A. Bensoussan, G. Da Prato, M. C. Delfour and S. K. Mitter, *Representation and Control of Infinite Dimensional Systems*, Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, 2007.

[15] A. Bensoussan, J. Frehse and P. Yam,*Mean Field Games and Mean Field Type Control Theory*, SpringerBriefs in Mathematics, Springer, New York, 2013.

[16] A. Bensoussan, T. Huang and M. Laurière, Mean field control and mean field game models with several populations, *Minimax Theory Appl.*, **3** (2018), 173–209.

[17] L. Briceño-Arias, D. Kalise, Z. Kobeissi, M. Laurière, Á. Mateos González and F. J. Silva, On the implementation of a primal-dual algorithm for second order time-dependent mean field games with local couplings, in *CEMRACS 2017–Numerical Methods for Stochastic Models: Control, Uncertainty Quantification, Mean-Field*, ESAIM Proc. Surveys, 65, EDP Sci., Les Ulis, 2019, 330–348.

[18] L. M. Briceño-Arias, D. Kalise and F. J. Silva, Proximal methods for stationary mean field games with local couplings, *SIAM J. Control Optim.*, **56** (2018), 801–836.

[19] H. Cao, X. Guo and M. Laurière, Connecting GANs, MFGs, and OT, preprint, `arXiv:2002.04112`.

[20] P. Cardaliaguet, *Notes on Mean Field Games*, 2013. Available from: `https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf`.

[21] P. Cardaliaguet and C.-A. Lehalle, Mean field game of controls and an application to trade crowding, *Math. Financ. Econ.*, **12** (2018), 335–363.

[22] E. Carlini and F. J. Silva. A fully discrete semi-Lagrangian scheme for a first order mean field game problem, *SIAM J. Numer. Anal.*, **52** (2014), 45–67.

[23] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications. I. Mean Field FBSDEs, Control, and Games*, Probability Theory and Stochastic Modelling, 83, Springer, Cham, 2018.

[24] R. Carmona, J.-P. Fouque and L.-H. Sun, Mean field games and systemic risk, *Commun. Math. Sci.*, **13** (2015), 911–933.

[25] R. Carmona, K. Hamidouche, M. Laurière and Z. Tan, Policy optimization for linear-quadratic zero-sum mean-field type games, *Proceedings of the IEEE Conference on Decision and Control*, Jeju, Korea, 2020.
[26] R. Carmona and M. Laurière, Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games I: The ergodic case, *SIAM J. Numer. Anal.*, **59** (2021), 1455–1485.
[27] R. Carmona and M. Laurière, Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games II: The finite horizon case, preprint, arXiv:1908.01613.
[28] R. Carmona, M. Laurière and Z. Tan, Linear-quadratic mean-field reinforcement learning: Convergence of policy gradient methods, preprint, arXiv:1910.04295.
[29] R. Carmona, M. Laurière and Z. Tan, Model-free mean-field reinforcement learning: Mean-field MDP and mean-field Q-learning, preprint, arXiv:1910.12802.
[30] A. Cherukuri, B. Gharesifard and J. Cortés, Saddle-point dynamics: Conditions for asymptotic stability of saddle points, *SIAM J. Control Optim.*, **55** (2017), 486–511.
[31] A. Cosso and H. Pham, Zero-sum stochastic differential games of generalized McKean–Vlasov type, *J. Math. Pures Appl. (9)*, **129** (2019), 180–212.
[32] C. Daskalakis and I. Panageas, The limit points of (optimistic) gradient descent in min-max optimization, *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018, 9256–9266. Available from: https://dl.acm.org/doi/pdf/10.5555/3327546.3327597.
[33] B. Djehiche and S. Hamadène, Optimal control and zero-sum stochastic differential game problems of mean-field type, *Appl. Math. Optim.*, **81** (2020), 933–960.
[34] B. Djehiche, A. Tcheukam and H. Tembine, Mean-field-type games in engineering, *AIMS Electronics and Electrical Engineering*, **1** (2017), 18–73.
[35] C. Domingo-Enrich, S. Jelassi, A. Mensch, G. M. Rotskoff and J. Bruna, A mean-field analysis of two-player zero-sum games, preprint, arXiv:2002.06277.
[36] R. Elie, T. Ichiba and M. Laurière, Large banking systems with default and recovery: A mean field game model, preprint, arXiv:2001.10206.
[37] R. Elie, J. Pérolat, M. Laurière, M. Geist and O. Pietquin, On the convergence of model free learning in mean field games, *Proceedings of the AAAI Conference on Artificial Intelligence*, **34** (2020), 7143–7150.
[38] M. Fazel, R. Ge, S. M. Kakade and M. Mesbahi, Global convergence of policy gradient methods for the linear quadratic regulator, preprint, arXiv:1801.05039.
[39] Z. Fu, Z. Yang, Y. Chen and Z. Wang, Actor-critic provably finds Nash equilibria of linear-quadratic mean-field games, preprint, arXiv:1910.07498.
[40] H. Gu, X. Guo, X. Wei and R. Xu, Mean-field controls with Q-learning for cooperative MARL: Convergence and complexity analysis, preprint, arXiv:2002.04131.
[41] X. Guo, A. Hu, R. Xu and J. Zhang, Learning mean-field games, *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019, 4967–4977.
[42] M. Huang, R. P. Malhamé and P. E. Caines, Large population stochastic dynamic games: Closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle, *Commun. Inf. Syst.*, **6** (2006), 221–251.
[43] C. Jin, P. Netrapalli and M. I. Jordan, What is local optimality in nonconvex-nonconcave minimax optimization?, preprint, arXiv:1902.00618.
[44] H. Kim, J. Park, M. Bennis, S.-L. Kim and M. Debbah, Mean-field game theoretic edge caching in ultra-dense networks, *IEEE Transactions on Vehicular Technology*, **69** (2019), 935–947.
[45] V. Kučera, The discrete Riccati equation of optimal control, *Kybernetika (Prague)*, **8** (1972), 430–447.
[46] J.-M. Lasry and P.-L. Lions, Mean field games, *Jpn. J. Math.*, **2** (2007), 229–260.
[47] Z. Liu, B. Wu and H. Lin, A mean field game approach to swarming robots control, 2018 Annual American Control Conference (ACC), Milwaukee, WI, 2018.
[48] T.-T. Lu and S.-H. Shiou, Inverses of 2×2 block matrices, *Comput. Math. Appl.*, **43** (2002), 119–129.
[49] E. Mazumdar, M. I. Jordan and S. S. Sastry, On finding local Nash equilibria (and only local Nash equilibria) in zero-sum continuous games, preprint, arXiv:1901.00838.

[50] F. Mériaux, V. Varma and S. Lasaulce, Mean field energy games in wireless networks, 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, 2012.
[51] M. Nouiehed, M. Sanjabi, T. Huang, J. D. Lee and M. Razaviyayn, Solving a class of non-convex min-max games using iterative first order methods, *Advances in Neural Information Processing Systems*, **32** (2019), 14934–14942.
[52] A. C. M. Ran and R. Vreugdenhil, Existence and comparison theorems for algebraic Riccati equations for continuous- and discrete-time systems, *Linear Algebra Appl.*, **99** (1988), 63–83.
[53] D. Shi, H. Gao, L. Wang, M. Pan, Z. Han and H. V. Poor, Mean field game guided deep reinforcement learning for task placement in cooperative multi-access edge computing, *IEEE Internet of Things Journal*, **7** (2020), 9330–9340.
[54] J. Sun, J. Yong and S. Zhang, Linear quadratic stochastic two-person zero-sum differential games in an infinite horizon, *ESAIM: Control Optim. Calc. Var.*, **22** (2016), 743–769.
[55] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, 2007.
[56] R. Xu, Zero-sum stochastic differential games of mean-field type and bsdes, *Proceedings of the 31st Chinese Control Conference*, (2012), 1651–1654.
[57] K. Zhang, Z. Yang and T. Basar, Policy optimization provably converges to Nash equilibria in zero-sum linear quadratic games, *Advances in Neural Information Processing Systems*, (2019) 11598–11610.

Received July 2020; revised July 2021; early access August 2021.

E-mail address: rcarmona@princeton.edu
E-mail address: kenzah@princeton.edu
E-mail address: lauriere@princeton.edu
E-mail address: zongjun.tan@princeton.edu