Mean field linear quadratic games with set up costs

Raffaele Pesenti, Dario Bauso

Mean field linear quadratic games with set up costs.

(Invited Paper)

Raffaele Pesenti
Dipartimento di Management
Università di Venezia - Ca' Foscari, Italy
Email: pesenti@unive.it

Dario Bauso
Dipartimento di Ingegneria Chimica,
Gestionale, Informatica e Meccanica
Università di Palermo, Italy
Email: dario.bauso@unipa.it

***Abstract*—This paper studies linear quadratic games with set up costs monotonic on the number of active players, namely, players whose action is non-zero. Such games arise naturally in joint replenishment inventory systems. Building upon a preliminary analysis of the properties of the best response strategies and Nash equilibria for the given game, the main contribution is the study of the same game under large population. Numerical illustrations are provided.**

I. Introduction

In this paper, we study linear quadratic games with set up costs monotonic on the number of active players, namely, players whose action is non-zero. Such games arise naturally in multi-retailer inventory application as shown in a previous work of the same authors [1].

In the first part of this paper, we analyze some properties of the best response strategies. In particular, we show that best response strategies are non-idle in the sense that a player never switches from being inactive to active for fixed behaviors of the other players (fixed set up costs). Non-idleness is used to derive an iterative procedure to compute Nash equilibria.

We then turn to consider large population games and in doing this we link our study to mean field games [2] [3] [4]. It turns out that most properties enjoyed by the game with finite players still hold when the number of players tends to infinity. This preliminary consideration allows us to claim that fixed points exist and that these are associated to mean field equilibria.

The paper is organized as follows. In Section II, we introduce the game. In Section III we analyze some properties of best response strategies. In Section IV, we discuss Nash equilibria. In Section V, we consider the game with large population and illustrate the mean field approach. In Section VI, we provide numerical illustrations and conclude in Section VII.

Notation. We denote by $P = \{1, 2, \ldots, n\}$ a set of n players. We use index i to refer to the generic ith player. Likewise, index $-i$ refers to all players other than i. We use $\mathbb{R}_+$ to denote the set of non-negative reals. Open and closed intervals between scalars a and b are denoted by $[a, b]$ and (a, b) respectively. We use $[0, T]$ to denote a finite horizon from 0 to T. Given a function of time $\phi(\cdot) : [0, T) \to \mathbb{R}$, we denote by $\phi(t)$ its value at time $t \in [0, T)$. We use $\phi[\xi](\cdot)$ to express the dependence of the function on a given parameter or function ξ.

II. Game definition

Each player $i \in P$ is characterized by the state variable $x_i(\cdot) \in \mathbb{R}$, the initial state $x_i^0 \in \mathbb{R}$, the measurable control $t \mapsto u_i(t)$, taking value, for all $t \in [0, T)$, in the set $\mathbb{R}$. The state variable evolves according to the dynamics

$$\begin{cases} \dot{x}_i(t) = u_i(t), & t \in [0, T) \\ x_i(0) = x_i^0 \end{cases} . \tag{1}$$

Let us also introduce the measurable opponents' control $t \mapsto u_{-i}(t)$, taking value, for all $t \in [0, T)$, in the set $\mathbb{R}^{n-1}$ and denote the sets for the measurable controls u and u_{-i} by

$$\begin{aligned} U_i &= \left\{ u_i : [0, T) \to \mathbb{R} \middle| u_i \text{ measurable} \right\}, \\ U_{-i} &= \left\{ u_{-i} : [0, T) \to \mathbb{R}^{n-1} \middle| u_{-i} \text{ measurable} \right\}. \end{aligned} \tag{2}$$

Let K, α, and β be given positive constants; $\delta : \mathbb{R} \to \{0, 1\}$ be defined as in (3) and $a : \mathbb{R}^{n-1} \to \mathbb{R}_+$ as in (4) where b is a constant greater than 0:

$$\delta(u_i(t)) = \begin{cases} 0 & \text{if } u_i(t) = 0, \\ 1 & \text{otherwise;} \end{cases} \tag{3}$$

$$a(u_{-i}(t)) = b + \frac{1}{n} \sum_{j \in P} \delta(u_j(t)). \tag{4}$$

The ith cost function is then

$$\begin{aligned} J_i(x_i^0, u_i, u_{-i}) \;\; = \;\; & \int_0^T \left(\tfrac{K\delta(u_i(t))}{a(u_{-i}(t))} + x_i(t)^2 \right. \\ & \left. + \alpha u_i(t)^2 \right) dt + \beta x_i(T)^2 \end{aligned} .$$

We say that a player is *active* at time t if its control $u_i(t)$ is non-null. Then, the above cost function implies that each player $i \in P$ pays for its state norm $x_i(t)^2$. In addition, if active, it pays both a fixed cost $\frac{K\delta(u_i(t))}{a(u_{-i}(t))}$ and a variable cost $\alpha u_i(t)^2$ for implementing its strategy. Observe that the fixed component of the cost is distributed among all the active players.

III. Properties of non-dominated strategies

Let the set of the *non-anticipating strategies* for the first player be

$$\begin{aligned} &M = \Big\{ \mu_i = \mu_i[x_i^0, \cdot] : U_{-i} \to U_i \Big| \\ &u_{-i}^a(s) = u_{-i}^b(s) \forall s \in [0,t] \Longrightarrow \\ &\mu_i[x_i^0, u_{-i}^a](s) = \mu_i[x_i^0, u_{-i}^b](s) \forall s \in [0,t], \\ &\forall u_{-i}^a, u_{-i}^b \in U_{-i}, \forall t \in [0,T) \Big\}. \end{aligned} \tag{5}$$

Hereafter, we consider only strategies $\mu_i[x_i^0, u_{-i}]$ such that

1) $\mu_i[x_i^0, u_{-i}](t) = 0$ or $sign(\mu_i[x_i^0, u_{-i}](t)) = -sign(x_i(t))$ where $x_i(t)$ is solution of
$$\begin{cases} \dot{x}_i(t) = \mu_i[x_i^0, u_{-i}](t), & t \in [0\,T) \\ x_i(0) = x_i^0 \end{cases}, \tag{6}$$
2) $\mu_i[x_i^0, u_{-i}]$ is piece-wise continuous.

There is no loss of generality in such a choice as, given the player i dynamics and cost, for no reason i would control its state so to increase its state norm.

We say that a strategy $\mu_i[x_i^0, u_{-i}]$ is *non-idle* if, for each interval $[t_1, t_2]$, $0 \le t_1 < t_2 \le T$ in which the set up cost $\frac{K}{a(u_{-i}(t))}$ is non-decreasing, $u_i(t) := \mu_i[x_i^0, u_{-i}](t) > 0$ for all $t_1 \le t \le t_1 + \Delta t$ and $u_i(t) = 0$ for all $t_1 + \Delta t < t \le t_2$, for some $0 \le \Delta t \le t_2 - t_1$. Then, a player i that implements a non-idle strategy, over the considered interval, is either always active or is always inactive or is first active and then inactive, but in no case it remains some time inactive before becoming active. Hereafter, we define *switching time instant*, the time in which a non-idle strategy $u(t)$ becomes non-active, i.e., the time $\inf\{t : u(t) = 0\}$.

The following two lemmas prove that a non-dominated strategy for a player i is non-idle and that the instantaneous set up cost paid by an active player cannot decrease over time.

Lemma 1: A strategy that is not non-idle is dominated.

Proof: Given a time interval $[t_1, t_2]$, $0 \le t_1 < t_2 \le T$ where $\frac{K}{a(u_{-i}(t))}$ does not decrease, consider two strategies $\mu_i^a[x_i^0, u_{-i}]$ and $\mu_i^b[x_i^0, u_{-i}]$ (see Fig. 1) such that $\mu_i^a[x_i^0, u_{-i}](t) = \mu_i^b[x_i^0, u_{-i}](t)$ for all $t \neq [t_1, t_2]$ and

$$\begin{cases} \mu_i^a[x_i^0, u_{-i}](t) = 0, & t \in [t_1, t_1 + \Delta t) \\ \mu_i^a[x_i^0, u_{-i}](t) \neq 0, & t \in [t_1 + \Delta t, t_2) \end{cases}$$

$$\begin{cases} \mu_i^b[x_i^0, u_{-i}](t) = \mu_i^a[x_i^0, u_{-i}](t + \Delta t), & t \in [t_1, t_2 - \Delta t) \\ \mu_i^b[x_i^0, u_{-i}](t) = 0, & t \in [t_2 - \Delta t, t_2) \end{cases}.$$

Denoting u_i^l the control where $u_i^l(t) := \mu_i^l[x_i^0, u_{-i}](t)$ and the label $l \in \{a, b\}$, we prove that control u_i^a is dominated by u_i^b.

To see this, let us denote by $x_i(t_1) = \int_0^{t_1} u_i^a(t)dt + x_i^0 = \int_0^{t_1} u_i^b(t)dt + x_i^0$, and $x_i(t_2) = \int_0^{t_2} u_i^a(t)dt + x_i^0 = \int_0^{t_2} u_i^b(t)dt + x_i^0$. In the following, we prove that

$$J_i(x_i^0, u_i^a) - J_i(x_i^0, u_i^b)) > 0. \tag{7}$$

Indeed, the costs induced by the two strategies are equal for $0 \le t \le t_1$ and $t_2 \le t \le T$, as in such interval the two strategies assume the same values and induce the same states for the player.

Then consider the interval $t_1 \le t \le t_2$, the cost paid by u_i^a is

$$\begin{aligned} &\int_{t_1}^{t_1+\Delta t} x_i(t_1)^2 dt + \int_{t_1+\Delta t}^{t_2} \frac{K}{a(u_{-i}(t))} dt + \\ &+ \int_{t_1+\Delta t}^{t_2} \alpha u_i^a(t)^2 dt + \int_{t_1+\Delta t}^{t_2} x_i(t)^2 dt. \end{aligned}$$

Differently, the cost paid by u_i^b is

$$\begin{aligned} &\int_{t_1}^{t_2-\Delta t} x_i(t)^2 dt + \int_{t_1}^{t_2-\Delta t} \frac{K}{a(u_{-i}(t))} dt + \\ &+ \int_{t_1}^{t_2-\Delta t} \alpha u_i^a(t + \Delta t)^2 dt + \int_{t_2-\Delta t}^{t_2} x_i(t_2)^2 dt. \end{aligned}$$

Now, note that $\int_{t_1+\Delta t}^{t_2} \frac{K}{a(u_{-i}(t))} dt \ge \int_{t_1}^{t_2-\Delta t} \frac{K}{a(u_{-i}(t))} dt$ since $\frac{K}{a(u_{-i}(t))}$ does not decrease for $t_1 \le t \le t_2$. In addition, observe that $\int_{t_1+\Delta t}^{t_2} u_i^a(t)^2 dt = \int_{t_1}^{t_2-\Delta t} u_i^a(t + \Delta t)^2 dt$, and $\int_{t_1+\Delta t}^{t_2} x_i(t)^2 dt = \int_{t_1}^{t_2-\Delta t} x_i(t)^2 dt$, then the inequality (7) holds true, as it becomes

$$J_i(x_i^0, u_i^a) - J_i(x_i^0, u_i^b)) \ge (x_i(t_1)^2 - x_i(t_2)^2)\Delta t > 0.$$

Hence, the lemma is proved. ■

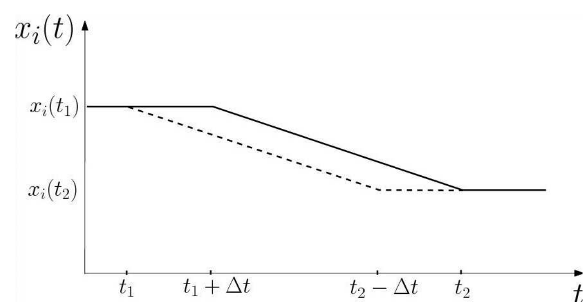

Fig. 1. Qualitative plot of the state evolution with strategies $\mu_i^a[x_i^0, u_{-i}]$ (solid) and $\mu_i^b[x_i^0, u_{-i}]$ (dashed) for $t \in [t_1, t_2]$ used in the proof of Lemma 1.

Lemma 2: If all the players play non-dominated strategies, $\frac{K}{a(u_{-i}(t))}$ does not decrease for all $i \in P$ and all $0 < t < T$.

Proof: The value $\frac{K}{a(u_{-i}(t))}$ decreases for some player i and some $0 < t < T$, if there is at least another player j that in t^j switches from being inactive to being active.

Let us first prove the result under the assumption that no more than one player can become active at each time instant. Then, there exists a value $t_1 \ge 0$, a value $\Delta t > 0$, and an interval $0 \le t_1 < t_1 + \Delta t = t^j < t_2 < T$, such that player j is first inactive and then active even if $\frac{K}{a(u_{-j}(t))}$ remains constant. Then, by Lemma 1, player j cannot be playing a non-dominated strategy.

Given the above argument, for $\frac{K}{a(u_{-i}(t))}$ to decrease for some player i, we must assume that a set S of players, with

$|S| \geq 2$, coordinates to switch from being inactive to being active at time $t^j > 0$. Even in this case, there exists a value $t_1 \geq 0$, a value Δt, and an interval $0 \leq t_1 < t_1 + \Delta t = t^j < T$, such that $\frac{K}{a(u_{-s}(t))}$ remains constant, for all $s \in S$. Following the same line of reasoning of Lemma 1, it is immediate to prove that strategies that coordinate the switch at time t_1 induce less costs for all the players in S and then they dominate the current strategies (that coordinate the switch at time t_j). We can conclude that the strategy that coordinates the switch at time t_j cannot be a non-dominated one. ■

We define *switching feedback strategy at* τ any control $\mathbf{u}_i[\tau]$ that satisfies:

$$\begin{aligned} \mathbf{u}_i[\tau](t) &:= \mu_i[x_i^0, u_{-i}](t) \\ &= \begin{cases} f(t,\tau)x_i(t) & \text{for } 0 \leq t \leq \tau \\ 0 & \text{for } \tau < t \leq T \end{cases}. \end{aligned} \tag{8}$$

In the hypotheses of the above two lemmas, the following corollary holds.

Corollary 1: For all τ such that $0 \leq \tau \leq T$ and for each player i there exists a unique non-dominated switching feedback strategy $\mathbf{u}_i[\tau]$ as in (8).

Proof: For all τ such that $0 \leq \tau \leq T$, the non-dominated strategy is

$$\mathbf{u}_i[\tau](t) = \begin{cases} \tilde{u}_i(t) & \text{for } 0 \leq t \leq \tau \\ 0 & \text{for } \tau < t \leq T \end{cases},$$

where $t \mapsto \tilde{u}_i(t)$ solves the problem below:

$$\begin{aligned} \tilde{u}_i &:= \arg\min\{\int_0^\tau (\frac{K}{a(u_{-i}(t))} + x_i(t)^2 + \alpha u_i(t)^2)dt \\ &\quad +((T-\tau)+\beta)x_i(\tau)^2\} \\ &= \arg\min\{\int_0^\tau (x_i(t)^2 + \alpha u_i(t)^2)dt \\ &\quad +((T-\tau)+\beta)x_i(\tau)^2\}. \end{aligned}$$

The equality holds as the value of $\frac{K}{a(u_{-i}(t))}$ is independent of $\tilde{u}_i(t)$. In addition, if $x_i^0 > 0$, the second problem, because of the quadratic structure of the costs, presents a unique optimal continuous solution of type $\tilde{u}_i(t) = f(t,\tau)x(t)$. The latter strategy is independent of the fixed cost and is different from zero for $0 \leq t \leq \tau$, as it can be directly verified explicitly solving the optimization problem. In this context note that this problem is a quadratic control problem that can be analytically solved using the maximum principle or a differential Riccati equation. ■

Hereafter, for any realization of u_{-i} and therefore $\frac{K}{a(u_{-i}(t))}$, we say that the best response strategy of player i is the switching feedback control at t_i^* defined as:

$$\begin{aligned} \mathbf{u}_i[t_i^*] &:= \arg\min_{\mathbf{u}_i[\tau]:0\leq\tau\leq T}\{J(x_0^i, \mathbf{u}_i[\tau], u_{-i})\} \\ &=: \mu_i^*[x_i^0, u_{-i}]. \end{aligned} \tag{9}$$

Note that a strategy solution of (9) always exists and is unique, as it can be verified analytically that $J(x_0^i, \mathbf{u}_i[\tau], u_{-i})$ is a continuous strictly convex function of τ. In case of multiple solutions, we observe that an immediate consequence of the above lemma is that the corresponding state trajectories do not intersect. Then, in the rest of the paper, we consider as best response strategy only the one that defines the state trajectory with minimal value for $0 \leq t \leq T$.

The next lemma relates the switching times of two different players.

Lemma 3: Given two players i and j, such that $x_i^0 \geq x_j^0 > 0$, if $\mathbf{u}_j[t_j^*]$ is a best response strategy for player j, then all the strategies of player i $\mathbf{u}_i[\tau]$ where $\tau < t_j^*$ are dominated.

Proof: The statement of this lemma can be directly verified by explicitly determining the values of $J(x_i^0, \mathbf{u}_i[\tau], u_{-i})$ and $J(x_j^0, \mathbf{u}_j[\tau], u_{-i})$ and observing that $J(x_i^0, \mathbf{u}_i[\tau], u_{-i})$ decreases for $\tau \in [0, t_j^*]$ as long as $J(x_j^0, \mathbf{u}_j[\tau], u_{-j})$ decreases in the same interval. The latter is true as $\mathbf{u}_j[t_j^*]$ is the best response strategy for player j. ■

The above lemma can be rephrased by saying that according to their best responses if player j is active then player i is active too.

The next theorem states under which condition a player active a time $t = 0$ becomes inactive in a following time instant. Specifically, it points out the dependence of the switching time instant of a non-dominated strategy on the value of the fixed cost.

Theorem 1: According to a non-dominated strategy, player i is active as long as the instantaneous set up cost satisfies the following condition

$$\frac{K\alpha}{a(u_{-i}(t))} \leq (((T-t)+\beta)x_i(t))^2. \tag{10}$$

When the above condition is satisfied, a non-dominated strategy is bounded as in (11), where $\gamma := -((T-t)+\beta)x_i(t)$ and $\Delta := (((T-t)+\beta)x_i(t))^2 - K\alpha/a(u_{-i}(t))$:

$$\frac{\gamma - \sqrt{\Delta}}{\alpha} \leq u_i(t) \leq \frac{\gamma + \sqrt{\Delta}}{\alpha}. \tag{11}$$

Proof: We analyze under which circumstances player i, active at time t, remains so for a further interval time $\Delta t > 0$. Then, let us look at interval $[t, t+\Delta t]$ and consider a *non-null strategy*, where $u(t) > 0$, and a *null strategy*, with $u(t) = 0$, for $t \in [t, t+\Delta t]$. Let us compare the cost to go from t to T induced by such strategies. The cost to go of the null strategy is

$$\int_t^{t+\Delta t} x_i(t)^2 d\tau + \int_{t+\Delta t}^T x_i(t)^2 d\tau + \beta x_i(T)^2.$$

Similarly, the cost to go of the non-null strategy is the one displayed below, with $\Delta x_i = \int_t^{t+\Delta t} u_i(\tau)d\tau$:

$$\begin{aligned} &\int_t^{t+\Delta t} \left(\frac{K}{a(u_{-i}(\tau))} + \alpha u_i(\tau)^2 + x_i(\tau)^2\right) d\tau \\ &+ \int_{t+\Delta t}^T (x_i(t) + \Delta x_i)^2(t)d\tau + \beta(x_i(t) + \Delta x_i)^2. \end{aligned}$$

Then, we compute the difference of the two costs for $\Delta t \to 0$, to obtain

$$\left(\frac{K}{a(u_{-i}(t))} + \alpha u_i(t)^2\right) dt + 2(T-t)x_i(t)dx_i + 2\beta x_i(t)dx_i.$$

Since $dx_i = u_i(t)dt$, after dividing by dt the latter can be rewritten as

$$\frac{K}{a(u_{-i}(t))} + \alpha u_i(t)^2 + 2(T-t)x_i(t)u_i(t) + 2\beta x_i(t)u_i(t).$$

Hence, the non-null strategy provides a lower cost than the null strategy, and therefore we would rather have $u_i(t) > 0$ in t, if and only if the above difference is non-positive, that is if $\alpha u_i^2(t) + 2((T-t)+\beta)x_i(t)u_i(t) + \frac{K}{a(u_{-i}(t))} \leq 0$. In turn, this last inequality holds if and only if conditions (10) and (11) are satisfied.

■

An immediate consequence of the above theorem is that a player is certainly never active if $K\alpha > (b+1-1/n)((T+\beta)x_i^0)^2$.

Lemma 3 also implies that if all the players $j \neq i$ play their best responses, and using $\mathbf{u}_{-i}[t^*_{-i}]$ to denote their set of best response strategies in compact form, then it holds

$$\frac{K\alpha}{a(\mathbf{u}_{-i}[t^*_{-i}](t^*_i))} = (((T-t^*_i)+\beta)x_i(t^*_i))^2. \quad (12)$$

IV. Nash equilibria

In this section, we show how to determine a set of Nash equilibria strategies for players in P under the assumption that $0 < x_1^0 \leq x_2^0 \leq \ldots \leq x_n^0$. To this end, we heavily exploit Lemma 3 to determine the best response of the players.

Preliminarily, for each player i let us define $\hat{K}_i := \frac{K}{b+\frac{n-i+1}{n}}$ and $\hat{t}_i$ the time instant, if exists, for which by applying a switching strategy $\mathbf{u}_i[\hat{t}_i]$ the following equality holds

$$\hat{K}_i\alpha = (((T-\hat{t}_i)+\beta)x_i(\hat{t}_i))^2. \quad (13)$$

If no $\hat{t}_i$ satisfies the above condition and furthermore if $\hat{K}_i\alpha > ((T+\beta)x_i^0)^2$, then we set $\hat{t}_i = 0$, otherwise if $\hat{K}_i\alpha < ((T+\beta)x_i^0)^2$ then we set $\hat{t}_i = T$. Note that $\mathbf{u}_i[\hat{t}_i]$ is the best response strategy for player i if $\frac{K}{a(u_{-i}(\hat{t}_i))} = \hat{K}_i$, that is, if at the switching time instant the only active players are the ones with state greater than or equal to $x_i(\hat{t}_i)$, or, that is the same, as the trajectories of best strategies cannot intersect, the only active players are the ones with initial state greater than or equal to x_i^0. In other words, $\hat{t}_i$ is the last time instant in which it is convenient for player i to remain active even if there are only other $n-i$ active players.

Lemma 3 implies that if all the players play their best responses, then strategy $\mathbf{u}_1[t_1^*]$ for player 1 must satisfy:

$$\frac{K}{a(\mathbf{u}_{-1}[t^*_{-1}](t))} = \begin{cases} \frac{K}{b+1} =: \hat{K}_1 & \text{if } 0 \leq t \leq t_1^* \\ 0 & \text{if } t_1^* < t \leq T \end{cases}.$$

From the latter condition, and invoking conditions (12)-(13), we can infer that $t_1^* = \hat{t}_1$ and also that player 1 has a unique non-dominated strategy $\mathbf{u}_1[t_1^*] = \mathbf{u}_1[\hat{t}_1]$.

Let us now consider the generic player $i > 1$. It holds

$$t_i^* = \max\{t^*_{i-1}, \hat{t}_i\}. \quad (14)$$

Indeed, Lemma 3 implies that player i must be active at least as long as player $i-1$ is active, hence $t_i^* \geq t^*_{i-1}$. Lemma 3 also implies that if $t_i^* > t^*_{i-1}$, then in t_i^* the only active players are the ones with state greater than or equal to $x_i(t_i^*)$, this in turn implies that t_i^* is either equal to t^*_{i-1} or equal to $\hat{t}_i$, that is that player i can consider only two strategies $\mathbf{u}_i[t^*_{i-1}]$ or $\mathbf{u}_i[\hat{t}_i]$. Finally, observe that player i chooses $\mathbf{u}_i[\hat{t}_i]$, if $\hat{t}_i > t^*_{i-1}$ because it is convenient for player i to remain active even if only other $n-i$ players are active after t^*_{i-1}.

The above argument points out how it is easy to practically determine Nash equilibrium strategies $\mathbf{u}_i[t_i^*]$ of the game under study. Indeed, the strategy of player 1 can be individuated without knowing the strategies of the other players, then, recursively, the strategy of player i can be derived only on the basis of the strategies of the previous $i-1$ players.

V. Large number of players

Let us now reformulate our game from a mean field perspective. To this end, let $m(x,t)$ be the distribution of the players' states at time t. Hereafter, we always assume that the support of $m(x,t)$ is a subset of $\mathbb{R}_+$. Let us also define the function $\tilde{a}(\cdot) : \mathbb{R}_+^{n-1} \mapsto \mathbb{R}$ as $\tilde{a}(x_{-i}) := b + \int_{x_i(t)}^{+\infty} dm(x,t)$. Then, we can rewrite $\hat{K}_i$ as

$$\hat{K}_i = \frac{K}{\tilde{a}(x_{-i})}. \quad (15)$$

The considerations over the state trajectories that precede Lemma 3 imply that $\int_{x_i(t)}^{+\infty} dm(x,t) = \int_{x_i^0}^{+\infty} dm(x,0)$ is invariant over time and, indeed, the arguments in Section IV allow to determine Nash equilibrium strategies only on the basis of the initial state distribution.

We are interested in determining the generic player best strategy in presence of a large number of players. Also, in the same context, we are interested in determining the evolution over time of the players' state cumulative distribution $Q(y,t) := \int_y^\infty dm(x,t)$.

A. Generic player i best strategy

Let us first consider the generic player i best strategy. The recursive equation (14) allows player i to determine the switching time instant t_i^* of its best strategy $\mathbf{u}_i[t_i^*]$ and hence to individuate the strategy itself. Unfortunately, equation (14) is of no practical use in presence of a large number of players as it would force player i to wait for the decision of all the players from 1 to $i-1$ before being able to compute t_i^*. For this reason, player i may decide to play an approximatively optimal strategy $\mathbf{u}_i[\tilde{t}_i^*]$ based on an estimate $\tilde{t}_i^*$ of t_i^*. In particular, we observe that we may rewrite equation (14) as

$$t_1^* = \max\{\hat{t}_i, \max_{j<i}\{\hat{t}_j\}\}.$$

Then, for any subset $S \subseteq \{1, 2, \ldots, i-1\}$, the value

$$\tilde{t}_i^* = \max\{\hat{t}_i, \max_{j \in S}\{\hat{t}_j\}\} \leq t_i^*$$

is an estimate, and in particular a lower bound, of the switching time instant t_i^*. Needless to say that the $\tilde{t}_i^*$ becomes a better and better estimate of t_i^*, and hence $\mathbf{u}_i[\tilde{t}_i^*]$ a better and better approximation of the best strategy $\mathbf{u}_i[t_i^*]$, as the subset S includes more and more elements of $\{1, 2, \dots, i-1\}$.

The above kind of approximate strategy requires that player i communicates with the players in S to acquire the values of $\hat{t}_j$. Player i can play a different approximate strategy that just needs the observation of the behavior of player $i-1$ as described in the following.

Player i remains active as long as $i-1$ is active. Then, at the switching time instant t_{i-1} of $i-1$, player i decides whether it is convenient to remain active or not and for how long. If all the players use such an approximate strategy, this approximation identifies the best strategy from the switching time instant of $i-1$ on. Indeed, from such time instant player i can determine its best strategy based on the number of active players: all the players from 1 to $i-1$ are not active any more, viceversa, all the players from $i+1$ to n remain active at least as long as i is active. Unfortunately, player i cannot play its best strategy until the switching time instant of $i-1$ as it cannot a priori know its value. As the optimal choice would be a strategy of type $\tilde{u}_i(t) = f(t)x(t)$, player i can approximate such a strategy, as an example fixing the value of $f(t)$ to a constant.

B. Evolution of the cumulative distribution

We now study how $Q(y,t) = \int_y^\infty dm(x,t)$ evolves over time. Specifically, as the trajectories of players with different initial states do not cross, it must satisfy the transport equation

$$\frac{\partial}{\partial t} Q(y,t) = -u(y,t) \frac{\partial}{\partial y} Q(y,t), \tag{16}$$

where $u(y,t)$ is the control applied at time t by a player with state $x(t) = y$.

As the best strategy of a player depends only on its initial state, we observe that, for each initial state x_0 and time instant t we can write $x(t) - x_0 = \int_0^t \tilde{u}(\tau) d\tau$, where $\tilde{u}$ is the best strategy of a player with initial state x_0. Then the solution of (16) is

$$Q(y,t) = Q(y - \int_0^t \tilde{u}(\tau) d\tau, 0),$$

as it can be directly verified computing the partial derivatives of $Q(y,t)$ and exploiting the fact that $\tilde{u}(t) = u(x(t), t)$.

The above results generalize to all the cases in which players choose strategies that depend only on the initial states. We also observe that the more the time to go $T - t$ gets closer to 0 the higher must be the state of a player for being convenient for the player to be active. Formally, there exists an increasing function $\lambda : [0,T] \to \mathbb{R}$ such that

$$u(y,t) = \begin{cases} 0 & \text{for } y \le \lambda(t) \\ f(t) y & \text{for } y > \lambda(t) \end{cases}.$$

Hence, we can rewrite $Q(y,t)$ as

$$Q(y,t) = \begin{cases} Q(y - \int_0^t \tilde{u}(\tau) d\tau, 0) & \text{for } 0 \le t \le \lambda^{-1}(y) \\ Q(y, \lambda^{-1}(y)) & \text{for } \lambda^{-1}(y) < t \le T \end{cases}.$$

n	T	α	β	K	$x_i(0)$	τ
200	20	20	1	1600	$[0, 150]$	$1, 1.5, \dots, T$

TABLE I
SIMULATIONS DATA.

VI. NUMERICAL ILLUSTRATIONS

In this section we provide numerical illustrations for a large number of players evolving according to system (1) and with simulations data as reported in Table I.

In particular, the number of players is $n = 200$ and the horizon is $T = 20$.

The parameters appearing in the cost (5) are set as follows: $\alpha = 20$, $\beta = 1$, and $K = 1600$. Initial states $x_i(0)$ for all i are uniformly distributed over the interval $[0, 150]$. We also discretize the set of possible switching times and so $\tau \in \{1, 1.5, \dots, T\}$.

The Algorithm used to numerically illustrate the players' behavior accepts the simulations data as input and returns the best response strategies $\mathbf{u}_i[t_i^*]$ as in (9) and the associated state distribution $dm(x,t)$.

The algorithm is designed as follows. First, we initialize the state by using the Matlab in-built functions *rand* to generate a realization of the random variable $x(0)$ and *sort* to reorder the agents for increasing states.

For every possible value of the switching time $\tau \in \{1, 1.5, \dots, T\}$, and for all players $i = 1, \dots, n$, we compute the optimal (we say optimal as for fixed τ the strategy $\mathbf{u}_i[\tau]$ is independent of the other players' behaviors) strategy $\mathbf{u}_i[\tau]$ as in (8).

To do this, we solve the following differential Riccati equation in the scalar variable $p(t)$ $t \in [0, \tau]$:

$$\dot{p}(t) = \frac{1}{2\alpha} p(t)^2 - 2, \quad p(\tau) = 2(T - \tau) + \beta.$$

The solution of the above ordinary differential equation with boundary value on final time is obtained using the Matlab in-built function *ode45* with step size 0.1. Function $f(t,\tau)$ appearing in (8) is then derived by setting $f(t,\tau) = -\frac{1}{2\alpha} p(t)$. As a result we have $\mathbf{u}_i[\tau](t) = -\frac{1}{2\alpha} p(t) x_i(t)$ for all $t \in [0, \tau]$.

We also compute the cost associated to each $\mathbf{u}_i[\tau]$ as illustrated in Fig. 2. From Fig. 2 one observes that the costs are convex and increasing on the initial state value $x_i(0)$ (higher curves correspond to higher $x_i(0)$). Also the minimum is increasing on the initial state value $x_i(0)$ and this is in accordance with the fact that the players' trajectories preserve their order through time as recalled repeatedly throughout the paper.

For every player $i = 1, \dots, n$, we then extract by brute force comparison, the strategy $\mathbf{u}_i[t_i^*]$ as in (9). Hence, we simulate the state evolution with $\mathbf{u}_i[t_i^*]$ and illustrate the results in Fig. 3. One can observe that for most of the players, especially those with a higher initial state, the switching time t_i^* is around 15. Players usually stop before reaching zero as expected in consequence of the presence of a fixed cost K in the cost

Algorithm

Input: Simulations data
Output: best response strategies $\mathbf{u}_i[t_i^*]$ (9) and associated state distribution $dm(x,t)$.

```
1 : Initialize state x(0) ← rand[0, 150],
2 : for τ = 1, 1.5, ..., T do
3 :     for player i = 1, ..., n do
4 :         compute u_i[τ] (8) and associated cost,
5 :     end for
6 : end for
7 : for player i = 1, ..., n do
8 :     extract u_i[t_i^*] as in (9);
        simulate state evolution with u_i[t_i^*];
        compute distribution dm(x, t).
9 : end for
```

function. A player with a state relatively close to zero at a time $t \approx T$ (t is approaching the end of the horizon T) will be inactive to avoid paying the fixed cost.

Finally, we compute the distribution $dm(x,t)$ at three different times, $t = 0, T/5, T$, and display the results in Fig. 4. One observes that at $t = 0$ (top) the players are uniformly random distributed over the interval $[0, 150]$. For $t = 4$ (approximately one fifth of the horizon, i.e., $t = T/5$) players are all distributed over the interval $[0, 60]$ (middle). For $t = 20$ (end of the horizon) all players have reached an equilibrium state close to but different from zero as evidenced by the peaks of $dm(x,T)$ (bottom).

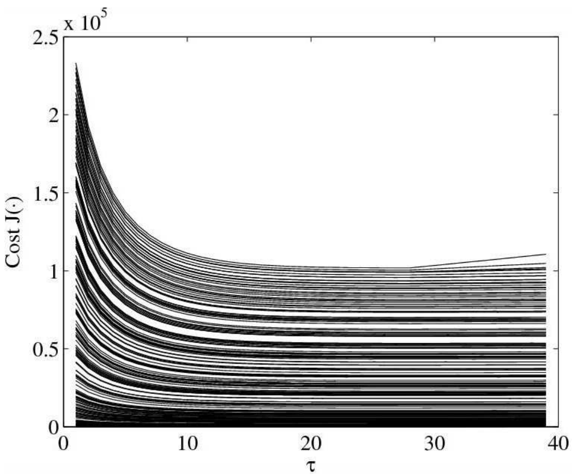

Fig. 2. Cost of switching feedback strategies at τ, $\mathbf{u}_i[\tau]$ as in (8), for different values of $\tau = 1, \ldots, T$.

VII. Conclusions

Inspired by joint replenishment inventory systems, we have introduced linear quadratic games with set up costs monotonic on the number of active players, namely, players whose action is non-zero. We have first analyzed the properties of the best response strategies and Nash equilibria for the given game. The obtained results are extended to the same game under large population.

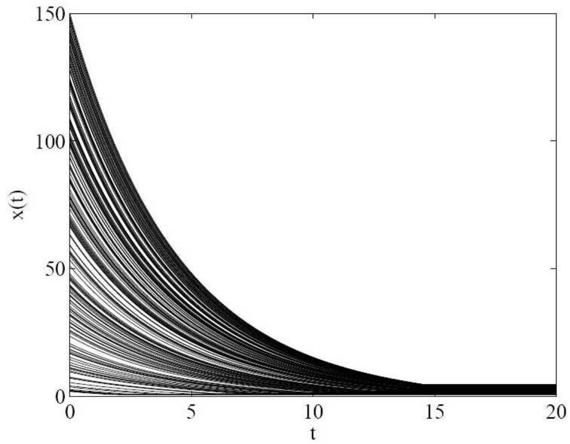

Fig. 3. Time plot of state $x(t)$ with best response strategies $\mathbf{u}_i[t_i^*]$ as in (9).

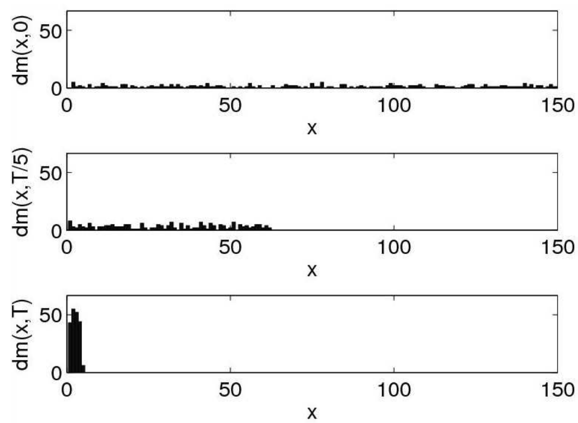

Fig. 4. Distribution $dm(x,t)$ for $t = 0$ (top), $t = T/5$ (middle), and $t = T$ (bottom) with best response strategies $\mathbf{u}_i[t_i^*]$ as in (9).

References

[1] D. Bauso, L. Giarrè, and R. Pesenti, "Consensus in noncooperative dynamic games: A multiretailer inventory application," *Trans. on Automatic Control*, vol. 53, no. 4, pp. 998–1003, 2008.

[2] M. Huang, P. Caines, and R. Malhamé, "Population cost-coupled lqg problems with non-uniform agents: Individual-mass behaviour and decentralized ϵ - nash equilibria," *Trans. on Automatic Control*, vol. 52, no. 9, pp. 1560–1571, 2007.

[3] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japanese Journal of Mathematics*, vol. 2, 2007.

[4] H. Tembine, Q. Zhu, and T. Basar, "Risk-sensitive mean field stochastic games," in *Proceedings of IFAC World Congress*, Milan, Italy, September 2011.