Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 50-1 (2017) 8957–8962

Linear Quadratic Stochastic Differential Games under Asymmetric Value of Information ⋆

Dipankar Maity and John S. Baras

Department of Electrical & Computer Engineering, Institute for Systems Research
University of Maryland, USA (e-mail: {dmaity, baras}@umd.edu).

Abstract: This paper considers a variant of two-player linear quadratic stochastic differential games. In this framework, none of the players has access to the state observations for all the time, which restricts the possibility of continuous feedback strategies. However, they can observe the state intermittently at discrete time instances by paying some finite cost. Having on demand costly measurements ensure that open-loop strategy is not the only strategy for this game. The individual cost functions for each player explicitly incorporate the value of information and the asymmetry that comes along with different costs of state observation for different players.
We study the structural properties of the Nash equilibrium for this particular class of problems when the cost of observation is finite and positive. We show that the game problem simplifies into two decoupled game problems: one for deciding the control strategies, and the other for deciding the observation acquisition times. The study also reveals that under two extreme cases -cost of observation being 0 or ∞- the strategies coincide with feedback and open-loop strategies respectively.

Keywords: game theory, linear-quadratic games, stochastic games, non-cooperative game, event-based game

1. INTRODUCTION

Game theory has been an active topic for research in control for its wide applicability in stochastic control, robust control; and it has been studied extensively by the community as can be found in Basar and Olsder (1995), James and Baras (1996), Engwerda (2005), Başar and Bernhard (2008), Fleming and Hernández-Hernández (2011) and many others. A differential stochastic game encompasses many aspects of a control problem such as optimality, stochasticity and filtering, and estimation; hence the results can reveal several properties related to those. Linear-quadratic-Gaussian is a subclass of such differential game problems that attain a closed form analytical solution for the Nash strategy as reported in Cruz Jr. and Chen (1971), Jacobson (1973), Weeren et al. (1999). The solution of linear-quadratic differential games are generally constructed by certain Riccati equation; for details, see Jacobson (1973), Weeren et al. (1999), and the references therein. Studies on the necessary and sufficient conditions for a strategy to be a Nash strategy for a linear-quadratic game can be found in the work of Foley and Schmitendorf (1971), and Bernhard (1979). Basar (1976) studied the uniqueness property of a Nash strategy. The work of Weeren et al. (1999) studies the asymptotic behavior of the Nash strategy over an infinite horizon.

Unlike the well perceived fact about linear control laws being optimal for a linear-quadratic-Gaussian problem, Basar (1974) provided a counterexample showing that the optimality is achieved by some nonlinear Nash strategy for a linear-quadratic game problem. A non-cooperative game is inherently a joint problem on the control and the decision making process and hence the solution relies on the knowledge of the behavior of the opponent.

In the vast majority of the past work in the community, the studied problems either assume that the state information is available to the players for all time or only the initial state information is available. The former situation results in a *feedback*-type Nash strategy whereas the latter exhibits an *openloop* Nash strategy. To the best of our knowledge, the problem of having multiple discrete state measurements for this problem class remained unaddressed. In this scenario, the players have less information about the state of the system since the measurements are available at only certain discrete time instances, not for all time; however, they have more information than merely having the knowledge of the initial state. In this work we address this linear-quadratic game problem under discrete measurements where the players are given the freedom to select their time instances to acquire the measurements of the state. Moreover, it is imposed that each such query about the state information requires some finite cost. This new framework introduces certain changes in the well known behavior of the Nash strategy since the feedback strategy is not plausible, and the open loop is not necessarily optimal. Given the fact that each observation

⋆ Research partially supported by ARO grants W911NF-14-1-0384 and W911NF-15-1-0646, and by National Science Foundation (NSF) grant CNS-1544787.

10.1016/j.ifacol.2017.08.1320

requires finite cost, the players must decide optimal time instances for observing the state. Therefore, the problem includes designing a sampling policy to measure the state and synthesizing a controller such that they constitute a Nash equilibrium for the game.

In this work, we assume that the sampling is done instantaneously and there is no delay or noise in communicating the sampled value to the controller. We consider asynchronous switching i.e. players can choose their switching policy irrespective of the policy chosen by their opponent. We also assume that whenever a player receives a sample, the opponent is notified about that but the opponent does not get the value of the sample.

In this study, we show that given a switching policy, there always exists a Nash strategy for controller synthesis and the controller is a dynamic controller that resets its value in an optimal way every time the switch is closed. The problem is decomposed into two decoupled subproblems for designing the switching policy and designing the controller. The studied game is asymmetric since the parameters associated with the players are different (e.g. cost per sample is different for different players) and that essentially leads to different strategy for them.

2. NOTATION

$x(t)$: state of the game, $\mathcal{C}_i$: controller of player-i, $\mathcal{S}_i$: switching policy of player-i, $\|a\|_B^2 = a'Ba$ for matrices a and b of proper dimensions, $\mathcal{T}_i(t)$: set of sampling times until t for player-i, $\mathcal{X}_i(t)$: set of sampled state values for player-i until t, $\mathcal{I}_i(t)$: total information available to player-i that includes $\mathcal{X}_i(t)$, $\mathcal{T}_1(t)$ and $\mathcal{T}_2(t)$. For any matrix M, $\Phi_M(t,s)$ denotes the associated state transition matrix.

3. PROBLEM FORMULATION

Let us consider the following stochastic linear differential game dynamics:

$$dx = (Ax + B_1u_1 + B_2u_2)dt + GdW_t \tag{1}$$

where $x \in \mathbb{R}^n$, $u_1 \in \mathbb{R}^{m_1}$, $u_2 \in \mathbb{R}^{m_2}$ and W_t is a p dimensional Wiener process noise, independent of the initial state $x(0)$, acting on the system. The associated quadratic cost is:

$$J(u_1,u_2) = \mathbb{E}\left[\int_0^T (x'Lx + u_1'R_1u_1 - u_2'R_2u_2)dt\right] \tag{2}$$

where $L, R_i \succ 0$. All the matrices A, B_i, L, R_i, G are time varying unless or otherwise mentioned in the paper.

The objective of player-1 (or player-2) is to minimize (or maximize) the cost functional (2) with the knowledge of $x(t)$ at some finite number of discrete time instances. Let us consider the schematic presented in Figure 1 where player-i has to design its controller $\mathcal{C}_i$ and the optimal switching policy $\mathcal{S}_i$. The switch $\mathcal{S}_i$ closes only for a time instance and opens immediately so that the controller gets the state value only at a single time instance. We assume there is no delay in the switching action or in the noise-less channel so that the controller $\mathcal{C}_i$ gets the state information precisely at the switching time instance.

Prior works on linear-quadratic differential games either consider the switches $\mathcal{S}_i$ are closed for all t or open for all $t > 0$. We study the characteristics of the game and the associated Nash strategy(s) for this special set up of the game. Maity and Baras (2016a) studies the problem when the switches $\mathcal{S}_1$ and $\mathcal{S}_2$ operate synchronously i.e. they samples the state at the same time instances.

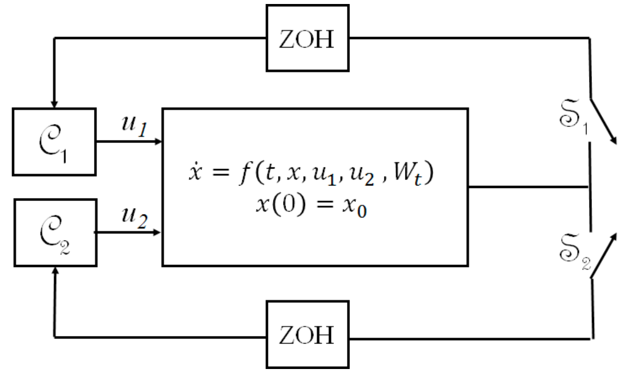

Fig. 1. Schematic of the game: $\mathcal{C}_i$ represents the controller (dynamic) of player-i and $\mathcal{S}_i$ implements a switch that samples the state x at some optimal instances. ZOH is a zero order hold circuit. The switches $\mathcal{S}_i$ are initially closed and they open at $t = 0^+$ so that each player has the knowledge of x_0.

In this work we assume that the state x is fully observable, however the study easily extends under partial observation framework along the lines of Maity et al. (2017). Moreover, the players are given the freedom to select their switching instances for sampling the state by incurring a finite cost.

Let $\mathcal{T}_i(t) = \{\tau_1^i, \tau_2^i, \cdots, \tau_{n_i(t)}^i\}$ be the set of selected time instances for closing the switch $\mathcal{S}_i$ of player-i till time t, where $\tau_k^i < \tau_{k+1}^i < \cdots < \tau_{n_i(t)}^i < t$. Let $\mathcal{T}(t) = \mathcal{T}_1(t) \cup \mathcal{T}_2(t)$ denote the set of all instances for observing the state. The state information available to player-i at time t is denoted as $\mathcal{X}_i(t) = \{x(\tau) \mid \tau \in \mathcal{T}_i(t)\}$. The total information available to player-i at any time instance t is $\mathcal{I}_i(t) = \mathcal{X}_i(t) \cup \mathcal{T}(t)$. However as mentioned, the information acquisition is not free and in order to construct $\mathcal{I}_i(t)$, player-i needs to pay $\lambda_i (> 0)$ for each sample of the state and $c_{ij} (> 0)$ for each element in $\mathcal{T}_j$ (player-j is the opponent of player-i). Thus, it should be noted that the cost function $J(u_1,u_2)$ is implicitly a function of the information set $\mathcal{I}_1$ and $\mathcal{I}_2$ since the strategies u_1 and u_2 are $\mathcal{I}_1$ and $\mathcal{I}_2$ measurable functions respectively ($J(u_1,u_2) \equiv J(u_1,u_2,\mathcal{I}_1,\mathcal{I}_2)$).

Therefore, player-1 (P1) needs to minimize:

$$\begin{aligned} J_1(u_1,u_2,\mathcal{I}_1,\mathcal{I}_2) =& \mathbb{E}\Big[\int_0^T (\|x\|_L^2 + \|u_1\|_{R_1}^2 - \|u_2\|_{R_2}^2)dt \\ &+ \lambda_1 N_1 + c_{12}N_2\Big], \end{aligned} \tag{3}$$

and player-2 (P2) should maximize:

$$\begin{aligned} J_2(u_1,u_2,\mathcal{I}_1,\mathcal{I}_2) =& \mathbb{E}\Big[\int_0^T (\|x\|_L^2 + \|u_1\|_{R_1}^2 - \|u_2\|_{R_2}^2)dt \\ &- \lambda_2 N_2 - c_{21}N_1\Big]. \end{aligned} \tag{4}$$

where $N_i = n_i(T)$ is the number of samples in $\mathcal{T}_i(T)$. In their respective cost functions, appropriate terms have been added to account for the cost of sampling (or could be thought as the cost of communication over the channel). It should be noted right away that the new cost functions do