2016 IEEE 55th Conference on Decision and Control (CDC)
ARIA Resort & Casino
December 12-14, 2016, Las Vegas, USA

Optimal Strategies for Stochastic Linear Quadratic Differential Games with Costly Information

Dipankar Maity and John S. Baras

***Abstract*— A two players stochastic differential game is considered with a given cost function. The players engage in a non-cooperative game where one tries to minimize and the other tries to maximize the cost. The players are given a dynamical system and their actions serve as the control inputs to the dynamical system. Their job is to control the state of this dynamical system to optimize the given objective function. We use the term "state of the game" to describe the state of this dynamical system. The challenge is that none of the players has access to the state of the game for all time, rather they can access the state intermittently and only after paying some information cost. Thus the cost structure is non-classical for a linear-quadratic game and it incorporates the value of information. We provide the Nash equilibrium strategy for the players under full state information access at no cost, as well as under costly state information access. The optimal instances for accessing the state information are also explicitly computed for the players.**

I. INTRODUCTION

Many facets of robust control, minimax stochastic control and general stochastic control problems have been revealed through the study of two players differential games [1], [2], [3], [4] [5]. To our best knowledge, differential games were first introduced with the zero-sum framework in [6]. Later, nonzero-sum differential games were introduced and studied in the works of Starr and Ho [7], [8]. Nash equilibrium strategies and their properties for different game frameworks have been studied for decades. Existence and uniqueness of such strategies are of paramount interests to studies like [9], [10] [11] and many others.

Among the classes of various differential games, linear-quadratic differential games is one of the most important and most studied class of problems. In this framework, the state of the game depends linearly on the strategies of the players and the optimization criteria for the players are quadratic. In most game problems, a closed form expression for the Nash strategy is not available and, in general, very difficult to calculate. However, for linear-quadratic games, the closed form expression for the Nash strategy is determined by solving certain coupled Riccati type equations [12], [10], [13]. Whereas [14], [9] studied the necessary and sufficient conditions for a strategy to be a Nash strategy, [11] addressed the problem of uniqueness of a Nash strategy. While most of the previous works address linear-quadratic games of fixed, finite time, [10] studies the asymptotic behavior of the Nash strategy when the duration of the game extends to infinity. Along with the studies of Nash strategies, Stackelberg strategies for nonzero-sum linear-quadratic games were investigated in [15]. Games where one of the players does not know the performance criterion of the opponent are best modeled and solved using the Stackelberg strategy concept [15].

[16] provided a counterexample showing the existence of a nonlinear Nash strategy for certain linear-quadratic games, startled the understanding that linear strategies are optimal for linear-quadratic games. In principle a non-cooperative game is a decision making problem and thus the information structures available to the players play a great role in determining the strategies for the players. In the class of linear-quadratic games, there are notions of open-loop, close-loop, feedback information patterns where the players can have the knowledge of the initial state of the game only, or all past and present states, or the present state only. This limited information determines the different structures of the corresponding Nash equilibrium strategies [17]. A very recent work [18] considered a differential game with the noise depending linearly on the state. The optimal strategy is a linear state feedback for the game [18], however it requires solving a stochastic Riccati-like equation for the players rather than a deterministic Riccati-type equation. The information pattern for this game is feedback type and hence the players can access the current state without paying any cost. The vast majority of past works makes the assumption that the state information is available to the players to construct their strategies.

In this work, we consider a nonclassical linear-quadratic differential game framework where we add a cost for acquiring state information. This new structure of the cost introduces notable changes in the well known structure of the Nash strategy for a similar game with no-cost information acquisition. We consider the game to have a continuous-time dynamics and the players can access state information only at discrete time instances and each such access query carries a finite cost. We impose the condition that the state access times for the players have to be the same and the state information will be made available to both of them at those time instances. At a first look, this seem to be a strong assumption but what this assumption technically means is that all the strategies of the players will be adapted to the same observational σ-field. Without this assumption, we will run into the problem of asymmetric information and the study of this problem is beyond the scope of this work. Some

Research partially supported by US Air Force Office of Scientific Research MURI grant FA9550-09-1-0538, by DARPA through ARO grant W911NF1410384, by National Science Foundation (NSF) grant CNS-1035655, and by NSF grant CNS-1544787.

The authors are with the Department of Electrical and Computer Engineering and The Institute for Systems Research, University of Maryland, College Park, USA. Email: dmaity@umd.edu, baras@umd.edu

preliminary results on this can be found in [19]. The analysis shows that the optimal state access times can be found offline by solving some finite-dimensional optimization problem. This entails the players to declare a unanimous decision of selecting state information acquisition times before the game starts. Therefore, we intend to study the following two facts of the game: first, how the players come to a consensus regarding the state accessing times and second, what will be their optimal strategies (the continuous controls that will drive the continuous dynamics of the game) for the game.

In the quadratic cost function, we take an integral cost and pointwise quadratic costs at several intermediate time instances. Pointwise cost at the final time (T) is widely studied in the framework of linear-quadratic control, however multiple pointwise costs allow the possibility of putting more emphasis on some discrete time instances. This unusual formulation of the cost changes the structure of the optimal strategy even with no-cost information access structure. The computation of the optimal strategy requires solving a Riccati-type equation semi-coupled with a linear ordinary differential equation ((8) and (9)).

Since the information acquisition is costly, the players will ask for information only at some optimal time instances. Similar to [19], the present study also shows that there is no direct relation between the optimal time instances of acquiring state information and the time instances of incurring pointwise cost.

We start our analysis by citing the optimal strategy for the players under no-cost information acquisition in Section III. Subsequently, we add the cost for information acquisition and solve the two decoupled optimization problems to describe the structure of the optimal strategy and the optimal way to construct the set containing the access times of the state information.

II. Problem Formulation

Let us consider a two player game $\mathcal{G}(x, J, u, v)$ as defined below.

Definition 2.1: A two players differential game $\mathcal{G}(x, J, u, v)$ consists of four components: $x(t)$: state of the underlying dynamical system at time t, $u(t), v(t)$: actions chosen by the players at time t, and J: the cost associated with the game. In a non-cooperative game set up, the objective of one player is to minimize the cost, whereas the other player maximizes the cost.

The state of the game (which means in this paper the state of the underlying stochastic dynamical system) obeys the following linear stochastic dynamics (1)

$$dx = Axdt + Budt + Cvdt + GdW_t \tag{1}$$

where W_t is an r dimensional standard Wiener process and the above dynamics should be treated in the sense of Ito. $\forall t$, $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{m_1}$ and $v(t) \in \mathbb{R}^{m_2}$. The model parameters $A, B, C,\ G$ are time variant, however, in order to maintain notational brevity, we omit the time argument.

The game is played for a finite time interval $[0, T]$ and the cost function is considered to be (2):

$$J(u,v) = \mathbf{E}\Big[\int_0^T (\|x\|_L^2+\|u\|_R^2-\|v\|_S^2)dt+\sum_{i=1}^{N}\|x(t_i)-r_i\|_{\alpha_i}^2\Big] \tag{2}$$

where $\|p\|_Q^2 = p'Qp$ for any two matrices p and Q of proper dimensions; r_i and α_i are some given fixed vectors and matrices that are the parameters of the cost function J. J consists of a running cost (integral term) and pointwise costs at some (N given) predefined time instances $0 \leq t_1 < t_2 < \cdots < t_N \leq T$.

The cost function has to be minimized for player-1 (P1) who selects the action u, while player-2 (P2) aims to maximize the cost by selecting the action v. In classical LQG control, there is (if any) only terminal state cost, however, the cost which we have considered is more general and it allows for penalization if the state at time t_i deviates from reference value r_i.

Each player has only limited and the same information about the state of the game and they can ask for the current state of the game by paying some cost. Let us denote the number of times state information is requested by the players up to time t be $n(t)$. The cost associated with these information acquisitions are $\lambda_1 n(t)$ and $\lambda_2 n(t)$ for P1 and P2 respectively. We add these costs to the cost function J. Hence P1 should minimize:

$$\begin{aligned} J_1(u,v) =& \mathbf{E}\Big[\int_0^T (\|x\|_L^2 + \|u\|_R^2 - \|v\|_S^2)dt \\ &+ \sum_{i=1}^{N}\|x(t_i) - r_i\|_{\alpha_i}^2 + \lambda_1 n(T)\Big], \end{aligned} \tag{3}$$

and P2 should maximize:

$$\begin{aligned} J_2(u,v) =& \mathbf{E}\Big[\int_0^T (\|x\|_L^2 + \|u\|_R^2 - \|v\|_S^2)dt \\ &+ \sum_{i=1}^{N}\|x(t_i) - r_i\|_{\alpha_i}^2 - \lambda_2 n(T)\Big]. \end{aligned} \tag{4}$$

$-\lambda_2 n(T)$ is added in (4) since P2 aims to maximize J. Though at this point we have two separate cost functions for each player, both the optimization problems boil down to minimizing and maximizing (2) once the players finalize the number of times the state will be accessed. The objective of the players is to jointly determine the time instances τ_k to construct their information set $\mathcal{I}(t) = \{x(\tau_k)\}_{k=1}^{n(t)}$ for all t. P1 selects the control u, as a function of this information set, to minimize $J(u, v)$. On the other hand, P2 constructs the strategy v, at time t, with the limited knowledge accumulated in $\mathcal{I}(t)$, to maximize $J(u, v)$ given in (2). The information set is ordered i.e. $\tau_k < \tau_{k+1}$ and non-anticipative i.e. for any t, $\tau_{n(t)} \leq t$. The set of admissible controls for P1 and P2 are adapted to the σ-field generated by the information set $\mathcal{I}(t)$.

III. Strategies with Costly Information

Before attempting the overall problem, let us first cite a theorem that gives the Nash Strategy for this game when the players do not have any cost for sampling the states, i.e. $\lambda_1 = \lambda_2 = 0$

Theorem 3.1 ([19]): The optimal strategies for the players are:

$$u^*(t) = -R^{-1}B'\left(\frac{\eta}{2} + Px\right) \quad (5)$$

$$v^*(t) = S^{-1}C'\left(\frac{\eta}{2} + Px\right). \quad (6)$$

The optimal cost is

$$J^* = \mathbf{E}[\|x(0)\|^2_{P(0)} + x(0)'\eta(0)] + \sum_{i=1}^{N} \|r_i\|^2_{\alpha_i} \quad (7)$$
$$+ \int_0^T \big(tr(PGG') + \frac{1}{4}\|\eta\|^2_{CS^{-1}C'} - \frac{1}{4}\|\eta\|^2_{BR^{-1}B'}\big)dt,$$

where

$$\dot{P} + A'P + PA + L + P(CS^{-1}C' - BR^{-1}B')P = \mathbf{0} \quad (8)$$
$$P(T) = \mathbf{0}$$
$$P(t_i^-) - P(t_i^+) = \alpha_i$$

and

$$\dot{\eta} = -\left(P(CS^{-1}C' - BR^{-1}B') + A'\right)\eta \quad (9)$$
$$\eta(T) = 0$$
$$\eta(t_i^-) - \eta(t_i^+) = -2\alpha_i r_i$$

Assumption 3.2: We assume $CS^{-1}C' - BR^{-1}B' \preceq 0$ $\forall t \in [0,T]$ in order to ensure existence and uniqueness of the solution of the Riccati equation (8) over $[0,T]$.

The following remarks follow directly from Theorem 3.1.

Remark 3.3 ([19]): *The optimal strategy consists of an open loop term to optimize the pointwise cost and a closed loop term to optimize both the pointwise cost and the integral cost.*

Remark 3.4 ([19]): *For the case when $t_N = T$, the jump and boundary conditions for (8) and (9) are given as follows:*

$$P(t_i^-) - P(t_i^+) = \alpha_i \ \forall 1 \le i < N$$

and $P(T) = \alpha_N$.

$$\eta(t_i^-) - \eta(t_i^+) = -2\alpha_i r_i \ \forall 1 \le i < N$$

and $\eta(T) = -2\alpha_N r_N$.

Remark 3.5 ([19]): *If the game parameters for both the players are the same i.e. $B = C$ and $R = S$, the game exhibits some interesting properties:*

- *The optimal strategies can be computed by solving two decoupled linear equations on P and η with the same boundary and jump conditions as in (8) and (9).*
- *The optimal cost in this case will be*

$$J^* = \mathbf{E}\big[\|x(0)\|^2_{\sum_{i=1}^{N}\|\Phi_A(t_i,0)\|^2_{\alpha_i} + \int_0^T \|\Phi_A(t,0)\|^2_L}$$
$$- 2x(0)'\sum_{i=1}^{N}\Phi_A(t_i,0)'\alpha_i r_i\big] + \sum_{i=1}^{N}\|r_i\|^2_{\alpha_i} \quad (10)$$
$$+ \int_0^T tr(PGG')dt$$

where (10) is derived by solving (8) and (9).

- *Any pair of strategies (u,v) of the form $(\gamma,-\gamma)$ will achieve the optimal cost. However, the pair $(-R^{-1}B'Px, R^{-1}B'Px)$ is the Nash equilibrium strategy.*

In Theorem 3.1, the explicit formulae for the optimal strategies of both the players require the knowledge of the state $x(t)$ for all time $t \in [0,T]$. Now we will investigate how the strategies change for both the players when they have state information only at finite number of time instances. Let us first attempt to solve the problem of finding the optimal strategies for both the players for an arbitrary interval $(t_0,t_1] \subseteq [0,T]$. In particular, we want to find the optimal strategies for the players in this interval but the strategies should not be continuous state feedback. The strategies may depend on the state information at discrete time instances and the strategies should not ask for future state information. In calculating the strategies, we still do not consider the state query cost at this point. In the subsequent sections we will remove this assumption and comment on the original problem. Making this assumption for this section makes our problem tractable for this initial step.

It can be shown [19] that:

$$J = \mathbf{E}\big[\|x(0)\|^2_{P(0)} + x(0)'\eta(0) + \sum_{i=1}^{N}\|r_i\|^2_{\alpha_i} \quad (11)$$
$$+ \int_0^T \big(\|u + R^{-1}B'(Px + \frac{\eta}{2})\|^2_R - \|v - S^{-1}C'(Px + \frac{\eta}{2})\|^2_S$$
$$- \frac{1}{4}\|\eta\|^2_{BR^{-1}B' - CS^{-1}C'} + tr(PGG')\big)dt\big]$$

From (11), we can divide the cost in two parts J_i, J_d. J_i is independent of the actions of the players and the other, J_d depends on the choice of u and v. Thus,

$$J_i = \mathbf{E}\big[\|x(0)\|^2_{P(0)} + x(0)'\eta(0) + \sum_{k=1}^{N}\|r_k\|^2_{\alpha_k} \quad (12)$$
$$+ \int_0^T \big(tr(PGG') - \frac{1}{4}\|\eta\|^2_{BR^{-1}B' - CS^{-1}C'}\big)dt\big]$$

and,

$$J_d(u,v) = \mathbf{E}\big[\int_0^T \big(\|u + R^{-1}B'(Px + \frac{\eta}{2})\|^2_R \quad (13)$$
$$- \|v - S^{-1}C'(Px + \frac{\eta}{2})\|^2_S\big)dt\big].$$

Since both players know the parameters of the game i.e. A, B, C, G, S and R, they can easily calculate the open-loop term. Thus, without loss of generality, it is sufficient for the players to optimize (14).

$$J_d(u,v) = \mathbf{E}\Big[\int_0^T \big(\|u + R^{-1}B'Px\|^2_R - \quad (14)$$
$$\|v - S^{-1}C'Px\|^2_S\big)dt\Big].$$

Let us denote $\tilde{R} = PBR^{-1}B'P$ and $\tilde{S} = PCS^{-1}C'P$. Without loss of generality, let the optimal controller for P1 be $u = -R^{-1}B'P\hat{x}_1$ and that of P2 be $v = S^{-1}C'P\hat{x}_2$, where $\hat{x}_i(t) = f(t, \mathcal{I}(t), \hat{x}_i(\tau)|_{\tau\in[0,t)})$ for some function f that needs to be determined. The purposes of $\hat{x}_1(t)$ and $\hat{x}_2(t)$ are to make optimal estimates of the state $x(t)$ by P1 and P2 respectively for all time t. Thus, rewriting J_d:

$$J_d(\hat{x}_1, \hat{x}_2) = \mathbf{E}[\int_0^T (\|x - \hat{x}_1\|^2_{\tilde{R}} - \|x - \hat{x}_2\|^2_{\tilde{S}})dt]. \quad (15)$$

With these strategies, the state of the game evolves as:

$$dx = (Ax + Bu_0 + Cv_0 - P^{-1}\tilde{R}\hat{x}_1 + P^{-1}\tilde{S}\hat{x}_2)dt + GdW_t. \quad (16)$$

Let us define $\Phi_Q(t,s)$ to be the solution of the following matrix differential equation.

$$\frac{d}{dt}\Phi_Q(t,s) = Q(t)\Phi_Q(t,s) \quad (17)$$
$$\Phi_Q(t,t) = \mathbf{I}, \quad (18)$$

for some matrix Q. Therefore, we can write the solution of (16) to be:

$$\begin{aligned} x(t) =& \Phi_A(t,t_0)x(t_0) + K^{t,t_0}[Bu_0 + Cv_0](t) \quad (19)\\ &- K^{t,t_0}[P^{-1}\tilde{R}\hat{x}_1 - P^{-1}\tilde{S}\hat{x}_2](t) + K_1^{t,t_0}[GW](t) \end{aligned}$$

where $K_1^{t_1,t_2}$ and K^{t_1,t_2} are linear operators defined as follows:

$$K^{t,t_0}[f](t) = \int_{t_0}^t \Phi_A(t,s)f(s)ds \quad (20)$$

$$K_1^{t,t_0}[fW](t) = \int_{t_0}^t \Phi_A(t,s)f(s)dW_s. \quad (21)$$

Let P1 select its strategy to be

$$\begin{aligned} \hat{x}_1(t) =& \Phi_A(t,t_0)\hat{x}_1(t_0) + K^{t,t_0}[Bu_0 + Cv_0 - P^{-1}\tilde{R}\hat{x}_1](t)\\ &+ f_1(t) \quad (22) \end{aligned}$$

Similarly, P2 selects its strategy in the following way

$$\begin{aligned} \hat{x}_2(t) =& \Phi_A(t,t_0)\hat{x}_2(t_0) + K^{t,t_0}[Bu_0 + Cv_0 + P^{-1}\tilde{S}\hat{x}_2](t)\\ &- f_2(t) \quad (23) \end{aligned}$$

where f_1 and f_2 are to be determined optimally. It should be noted at this point that these strategies do not have any special structures since there is no constraint on choosing f_1 and f_2. Writing the strategies in that form only provides some advantages in the analysis and computations. f_1 and f_2 are possibly piecewise continuous functions. There may be jump discontinuities in them in order to reset the values of $\hat{x}_1$ and $\hat{x}_2$ at the instances when the value of $x(t)$ is available. Using (22) and (23) for any interval $(t_0, t_1]$, we get

$$\begin{aligned} &\mathbf{E}[\int_{t_0}^{t_1} \|x - \hat{x}_1\|^2_{\tilde{R}} - \|x - \hat{x}_2\|^2_{\tilde{S}}dt]\\ =&\mathbf{E}[\int_{t_0}^{t_1} (\|K^{t,t_0}[P^{-1}\tilde{S}\hat{x}_2](t) - f_1(t) + \phi_A(t,t_0)e_1(t_0)\|^2_{\tilde{R}} -\\ &\quad \|K^{t,t_0}[P^{-1}\tilde{R}\hat{x}_1](t) - f_2(t) - \phi_A(t,t_0)e_2(t_0)\|^2_{\tilde{S}})dt\\ &+ \int_{t_0}^{t_1} \|K_1^{t,t_0}[GW](t)\|^2_{\tilde{R}-\tilde{S}}dt] \quad (24) \end{aligned}$$

where $e_1 = x - \hat{x}_1$ and $e_2 = x - \hat{x}_2$. $\hat{x}_1$ depends linearly on f_1 and hence $K^{t,t_0}[P^{-1}\tilde{R}\hat{x}_1](\cdot)$ can be written as an affine functional of f_1. Let us denote $\tilde{A}_R = A - P^{-1}\tilde{R}$, $\Delta_R\Phi(t,t_0) = \Phi_A(t,t_0) - \Phi_{\tilde{A}_R}(t,t_0)$, $\tilde{A}_S = A + P^{-1}\tilde{S}$, $\Delta_S\Phi(t,t_0) = \Phi_{\tilde{A}_S}(t,t_0) - \Phi_A(t,t_0)$. We define two new linear operators $K_R^{t,t_0}[\cdot]$ and $K_S^{t,t_0}[\cdot]$ as follows:

$$K_R^{t,t_0}[f](t) = \int_{t_0}^t \Phi_{\tilde{A}_R}(t,s)f(s)ds,$$
$$K_S^{t,t_0}[f](t) = \int_{t_0}^t \Phi_{\tilde{A}_S}(t,s)f(s)ds.$$

$\forall t \in (t_0, t_1]$, we define $K^{t,t_0}[P^{-1}\tilde{R}\hat{x}_1](t) = L_R[f_1](t)$. It can be shown that $L_R[\cdot]$ is an affine functional of the form:

$$\begin{aligned} L_R[f_1](t) =& \Delta_R\Phi_A(t,t_0)\left(\hat{x}_1(t_0^+) - f_1(t_0^+)\right)\\ &+ (K^{t,t_0} - K_R^{t,t_0})[Bu_0 + Cv_0](t)\\ &+ K_R^{t,t_0}[P^{-1}\tilde{R}f_1](t). \quad (25) \end{aligned}$$

Similarly it can also be shown that

$$\begin{aligned} L_S[f_2](t) &\triangleq K^{t,t_0}[P^{-1}\tilde{S}\hat{x}_2](t)\\ &= \Delta_S\Phi_A(t,t_0)\left(\hat{x}_2(t_0^+) + f_2(t_0^+)\right)\\ &+ (K_S^{t,t_0} - K^{t,t_0})[Bu_0 + Cv_0](t)\\ &- K_S^{t,t_0}[P^{-1}\tilde{S}f_2](t). \quad (26) \end{aligned}$$

Therefore, optimal f_1 and f_2 are chosen in order to optimize

$$\begin{aligned} \mathcal{J} =& \mathbf{E}[\int_{t_0}^{t_1} \left(\|L_S[f_2](t) - f_1(t) + \phi_A(t,t_0)e_1(t_0)\|^2_{\tilde{R}}\right.\\ &\left. - \|L_R[f_1](t) - f_2(t) - \phi_A(t,t_0)e_2(t_0)\|^2_{\tilde{S}}\right)dt]. \quad (27) \end{aligned}$$

Theorem 3.6: *$\mathcal{J}$ has a unique saddle point at f_1^* and f_2^* such that:*

$$\mathcal{J}(f_1^*, f_2) \le \mathcal{J}(f_1^*, f_2^*) \le \mathcal{J}(f_1, f_2^*) \quad (28)$$

where, for a given interval $(t_0, t_1]$[1], the optimal f_1^* and f_2^* satisfy the following differential equations:

$$\dot{f}_1^* = Af_1^* + P^{-1}\tilde{S}\zeta \quad (29)$$
$$f_1^*(t_0^+) = x(t_0) - \hat{x}_1(t_0)$$
$$\dot{\zeta} = \tilde{A}_S\zeta + Bu_0 + Cv_0 - P^{-1}\tilde{R}\hat{x}_1 \quad (30)$$
$$\zeta(t_0^+) = x(t_0)$$

$$\dot{f}_2^* = Af_2^* - P^{-1}\tilde{R}\xi \quad (31)$$
$$f_2^*(t_0^+) = \hat{x}_2(t_0) - x(t_0)$$
$$\dot{\xi} = \tilde{A}_R\xi + Bu_0 + Cv_0 + P^{-1}\tilde{S}\hat{x}_2 \quad (32)$$
$$\xi(t_0^+) = x(t_0)$$

Proof: The proof is provided in the Appendix V-A. ■

Remark 3.7: *The optimal $\hat{x}_1$ satisfies the following system of equations for the time interval $(t_0, t_1]$:*

$$\dot{\hat{x}}_1 = (A - P^{-1}\tilde{R})\hat{x}_1 + Bu_0 + Cv_0 + P^{-1}\tilde{S}\zeta \quad (33)$$
$$\dot{\zeta} = (A + P^{-1}\tilde{S})\zeta + Bu_0 + Cv_0 - P^{-1}\tilde{R}\hat{x}_1 \quad (34)$$
$$\hat{x}_1(t_0^+) = \zeta(t_0^+) = x(t_0).$$

[1]In this paper we use the following convention to define derivatives at enclosed boundary points: $\dot{f}(t_1) = \lim_{t\to t_1} \dot{f}(t)$ s.t. $t \in (t_0, t_1]$

Comparing (16) *and* (33), ζ *serves as the estimate of* $\hat{x}_2$ *for the evolution of* $\hat{x}_1$.

Remark 3.8: *The optimal* $\hat{x}_2$ *satisfies the following system of equations for the time interval* $(t_0, t_1]$*:*

$$\dot{\hat{x}}_2 = (A + P^{-1}\tilde{S})\hat{x}_2 + Bu_0 + Cv_0 - P^{-1}\tilde{R}\xi \quad (35)$$
$$\dot{\xi} = (A - P^{-1}\tilde{R})\xi + Bu_0 + Cv_0 + P^{-1}\tilde{S}\hat{x}_2 \quad (36)$$
$$\hat{x}_2(t_0^+) = \xi(t_0^+) = x(t_0).$$

Comparing (16) *and* (35), ξ *serves as the estimate of* $\hat{x}_1$ *for the evolution of* $\hat{x}_2$.

With this optimal strategy, the cost incurred is:

$$\mathbf{E}\left[\int_{t_0}^{t_1} (\|x - \hat{x}_1\|^2_{\tilde{R}} - \|x - \hat{x}_2\|^2_{\tilde{S}})\ dt\right] = \mathbf{E}\left[\int_{t_0}^{t_1} \|K_1^{t,t_0}[GW](t)\|^2_{\tilde{R}-\tilde{S}} dt\right].$$

So far we have been able to find the Nash strategy for the players in an arbitrary interval $(t_0, t_1]$. Thus, if somehow the players come to an agreement about the state information acquisition times $\{\tau_i\}_{i=1}^{n_T}$, they can find their strategies for the intervals $(\tau_i, \tau_{i+1}]$ where $\tau_0 = t_0$ and $\tau_{n_T+1} = T$. Since the strategies on different intervals are independent of each other, the strategies for the entire time horizon $[0, T]$ can be constructed by concatenating the individual strategies over the intervals $(\tau_i, \tau_{i+1}]$. The question that remains to be answered is how the players come to an agreement about the state information acquisition times. Let us analyse this from the point of view of both players. For P1, it is of paramount interest to optimize (3) which becomes (37) when the optimal f_1^* and f_2^* are used:

For P1:

$$\min_{n_1(T),\tau_1^1,\cdots,\tau^1_{n_1(T)}} \sum_{i=0}^{n_1(T)+1} \mathbf{E}\Big[\int_{\tau_i^1}^{\tau_{i+1}^1} \|K_1^{t,\tau_i^1}[GW](t)\|^2_{\tilde{R}-\tilde{S}} dt + \lambda_1 n_1(T)\Big]. \quad (37)$$

Whereas, P2 will seek to optimize the following function:

For P2:

$$\max_{n_2(T),\tau_1^2,\cdots,\tau^2_{n_2(T)}} \sum_{i=0}^{n_2(T)+1} \mathbf{E}\Big[\int_{\tau_i^2}^{\tau_{i+1}^2} \|K_1^{t,\tau_i^2} GW\|^2_{\tilde{R}-\tilde{S}} dt - \lambda_2 n_2(T)\Big] \quad (38)$$

where $\tau_0^1 = \tau_0^2 = 0$ and $\tau^2_{n_2(T)+1} = \tau^1_{n_1(T)+1} = T$.

Under the assumption that $\tilde{R}(t) - \tilde{S}(t) \succeq 0$, $\forall t \in [0, T]$, in order to ensure that the solution of the Riccati equation (8) is well defined, one can easily find out that for P2, the optimal choice would be to never access the state information. However, this is not the case for P1 and the choice for P1 depends on the value of λ_1 and the game parameters.

Remark 3.9: *If the game parameters for both the players satisfy the condition* $CS^{-1}C' = BR^{-1}B'$*, both the optimization problems for the players have the same solution, and that solution does not ask for any state information except* $x(t_0)$.

Remark 3.10: *The optimal choice of information set and the optimal strategy selection are two decoupled problems for each player.*

Let us denote the optimal value of the optimization problem (37) to be c_1^* and that of (38) to be c_2^*. If the jointly selected time instances are $\{\tau_1, \tau_2, \cdots, \tau_{N_T}\}$, then both the players will try to keep the cost incurred by this choice to be as close as possible to their optimal values. Therefore the following function needs to be minimized:

$$H(N_T, \tau_1, \cdots \tau_{N_T}) = \quad (39)$$
$$\Big\|\sum_{i=0}^{N_T} \int_{\tau_i}^{\tau_{i+1}} \mathbf{E}\|K^{t,\tau_i}[GW]\|^2_{\tilde{R}-\tilde{S}} dt + \lambda_1 N_T - c_1^*\Big\|^2 + \Big\|\sum_{i=0}^{N_T} \int_{\tau_i}^{\tau_{i+1}} \mathbf{E}\|K^{t,\tau_i}[GW]\|^2_{\tilde{R}-\tilde{S}} dt - \lambda_2 N_T - c_2^*\Big\|^2.$$

In (39) the first term tries to keep the cost of P1 close to the optimal value c_1^* whereas the second term tries to keep the cost of P2 close to its optimal value c_2^*.

In principle, there can be other functions that can model the same trade-off between the players, however in this paper we will consider (39) for the remainder of this paper.

From the properties of the Wiener process one can obtain,

$$\mathbf{E}[\|K^{t,\tau_i} GW\|^2_{\tilde{R}-\tilde{S}}] = \int_{\tau_i}^{t} tr(\|\Phi_A(t,s)G(s)\|^2_{\tilde{R}(s)-\tilde{S}(s)}) ds. \quad (40)$$

Let us denote $\sum_{i=0}^{N_T} \int_{\tau_i}^{\tau_{i+1}} \mathbf{E}\|K^{t,\tau_i}[GW]\|^2_{\tilde{R}-\tilde{S}} dt = C(N_T, \tau_1, \cdots, \tau_{N_T})$. Therefore, $H(N_T, \tau_1, \cdots, \tau_{N_T}) = \|C + \lambda_1 N_T - c_1^*\|^2 + \|C - \lambda_2 N_T - c_2^*\|^2$.

For a fixed N_T, to select optimal τ_k, we seek: $\frac{\partial H}{\partial \tau_i} = 0$. Therefore the necessary conditions are:

$$(2C + (\lambda_1 - \lambda_2)N_T - c_1^* - c_2^*)\frac{\partial C}{\partial \tau_i} = 0. \quad (41)$$

For all $i = 1, \cdots, N_T$, $\frac{\partial C}{\partial \tau_i} = 0$ implies

$$\int_{\tau_{i-1}}^{\tau_i} tr(\|\Phi_A(\tau_i,t)G(t)\|^2_{\tilde{R}(t)-\tilde{S}(t)}) dt = \int_{\tau_i}^{\tau_{i+1}} tr(\|\Phi_A(t,\tau_i)G(\tau_i)\|^2_{\tilde{R}(\tau_i)-\tilde{S}(\tau_i)}) dt, \quad (42)$$

which needs to be satisfied or $(2C + (\lambda_1 - \lambda_2)N_T - c_1^* - c_2^*) = 0$ admits a solution.

Claim 3.11: *For a fixed* N_T, $2C(N_T, \tau_1, \cdots, \tau_{N_T}) + (\lambda_1 - \lambda_2)N_T - c_1^* - c_2^* = 0$ *has a solution.*

The proof of the above claim follows directly from the fact that $C(N_T, \tau_1, \cdots, \tau_{N_T})$ is a continuous function of τ_i and the maximum and minimum values of C are $c_2^* + \lambda_2 N_T$ and $c_1^* - \lambda_1 N_T$ respectively. Therefore, there is a point in the space where the function attains a value equal to the average of its maximum and minimum values. Moreover, it is straightforward to show that $\{\tau_i\}$ satisfying $2C(N_T, \tau_1, \cdots, \tau_{N_T}) + (\lambda_1 - \lambda_2)N_T - c_1^* - c_2^* = 0$ is optimal. After this point, H will be a function of an integer variable N_T and its optimization can be done easily. Under this choice of sampling instances, the costs incurred by P1 and P2 are respectively $\frac{1}{2}[(c_1^* + c_2^*) + (\lambda_1 + \lambda_2)N_T]$ and $\frac{1}{2}[(c_1^* + c_2^*) - (\lambda_1 + \lambda_2)N_T]$.

Remark 3.12: *The game parameter G has influence in determining $\{\tau_k\}$, but the optimal strategies for the players do not rely on G. For a deterministic game ($G \equiv 0$) the results imply that the players do not need any more information other than $x(0)$. In fact, the deterministic framework results in the same strategy and state information access instances as the symmetric (i.e. $B = C$, $S = R$) game.*

Remark 3.13: *The set of state information acquisition times $\{\tau_k\}$ selected by the players is independent of the time instances of incurring the pointwise cost i.e. $\{t_i\}$.*

IV. Conclusions

In this work we have considered a two players stochastic linear-quadratic differential game. The Nash equilibrium strategy for both the players under full state information has been derived. With costly noise-free state information, we have derived the optimal instances for obtaining the state information for the players and the optimal strategies for selecting the actions u and v.

This framework can be easily extended to related problems such as the cost function being linear-exponential-quadratic or quadratic with exponential forgetting factor. The results are similar with minute difference and thus omitted from the paper due to limited space.

A valid question to ask is what happens when the players can independently choose their own set of triggering instances. The answer to this question is beyond the scope of this paper and will be considered as a possible future work.

V. Appendix

A. Proof of Theorem 3.6:

The Gateaux differential of the functional $\mathcal{J}$ is:

$$\delta\mathcal{J}[f_1, f_2](h_1, h_2) = \lim_{a\to 0} \frac{\mathcal{J}(f_1 + ah_1, f_2 + ah_2) - \mathcal{J}(f_1, f_2)}{a} \tag{43}$$

where the notation $\mathcal{J}[f_1, f_2](h_1, h_2)$ means the Gateaux differential of $\mathcal{J}$ evaluated at the point (f_1, f_2) in the direction (h_1, h_2). Note that $\mathcal{J}[f_1, f_2](\cdot, \cdot)$ is a linear functional parameterized by f_1 and f_2. Therefore,

$$\begin{aligned}&\frac{1}{2}\delta\mathcal{J}[f_1, f_2](h_1, h_2) = \\ &\mathbf{E}\Big[\int_{t_0}^{t_1} \big(\langle \Delta_S\Phi(t, t_0)h_2(t_0) - K_S^{t,t_0}[P^{-1}\tilde{S}h_2](t) - h_1(t), \\ &\quad L_S[f_2](t) - f_1(t) + \Phi_A(t, t_0)e_1(t_0)\rangle_{\tilde{R}} - \\ &\quad \langle K_R^{t,t_0}[P^{-1}\tilde{R}h_1](t) - \Delta_R\Phi(t, t_0)h_1(t_0) - h_2(t), \\ &\quad L_R[f_1](t) - f_2(t) - \Phi_A(t, t_0)e_2(t_0)\rangle_{\tilde{S}}\big)dt\Big]\end{aligned} \tag{44}$$

where $\langle a, b\rangle_C = a'Cb$ and $a, b,\ C$ are matrices (or vectors) of compatible dimensions.

After few steps it can be shown that:

$$\frac{1}{2}\delta\mathcal{J}[f_1, f_2](h_1, h_2) = \mathbf{E}[I_1 + I_2 + I_3 + I_4] \tag{45}$$

where

$$\begin{aligned}I_1 = -\int_{t_0}^{t_1} \Big(&\langle L_S[f_2](t) - f_1(t) + \Phi_A(t, t_0)e_1(t_0) \\ &+ P^{-1}(t)\int_t^{t_1} \Phi_{\tilde{A}_R}(s, t)'\tilde{S}(s)\big(L_R[f_1](s) - f_2(s) \\ &- \Phi_A(s, t_0)e_2(t_0)\big)ds,\ h_1\rangle_{\tilde{R}}\Big)dt\end{aligned} \tag{46}$$

$$\begin{aligned}&I_2 = \\ &\int_{t_0}^{t_1} \langle L_R[f_1] - f_2 - \Phi_A(t, t_0)e_2(t_0), \Delta_R\Phi(t, t_0)h_1(t_0)\rangle_{\tilde{S}}dt\end{aligned} \tag{47}$$

$$\begin{aligned}I_3 = \int_{t_0}^{t_1} \Big(&\langle L_R[f_1](t) - f_2(t) - \Phi_A(t, t_0)e_2(t_0) \\ &- P^{-1}(t)\int_t^{t_1} \Phi_{\tilde{A}_S}(s, t)'\tilde{R}(s)\big(L_S[f_2](s) - f_1(s) \\ &+ \Phi_A(s, t_0)e_1(t_0)\big)ds,\ h_2\rangle_{\tilde{S}}\Big)dt\end{aligned} \tag{48}$$

$$\begin{aligned}&I_4 = \\ &\int_{t_0}^{t_1} \langle L_S[f_2] - f_1 + \Phi_A(t, t_0)e_1(t_0), \Delta_S\Phi(t, t_0)h_2(t_0)\rangle_{\tilde{R}}dt\end{aligned} \tag{49}$$

For $\mathcal{I}(t)$ adapted f_1^* and f_2^* to be a stationary point for the functional $\mathcal{J}$, we need $\delta\mathcal{J}[f_1^*, f_2^*](h_1, h_2) = 0$ for all piecewise continuous $\mathcal{I}(t)$ measurable functions h_1 and h_2. From (45), (46), (47), (48) and (49) we obtain the necessary conditions to be held for all $t \in (t_0, t_1)$:

$$L_S[f_2](t) - f_1(t) + \Phi_A(t, t_0)\mathbf{E}[e_1(t_0) \mid \mathcal{I}(t)] = 0, \tag{50}$$

$$L_R[f_1](t) - f_2(t) - \Phi_A(t, t_0)\mathbf{E}[e_2(t_0) \mid \mathcal{I}(t)] = 0. \tag{51}$$

Evaluating (50) at t_0^+ and using the fact that $L_S[f_2](t_0^+) = 0$, we obtain:

$$f_1(t_0^+) = \mathbf{E}[x(t_0) \mid \mathcal{I}(t)] - \hat{x}_1(t_0). \tag{52}$$

If $x(t_0) \in \mathcal{I}(t)$, then $\mathbf{E}[x(t_0) \mid \mathcal{I}(t)] = x(t_0)$ i.e. the best possible estimate. This reflects the fact that t_0 is the time instance when the state of the game $x(t_0)$ must be made available. This fact along with (22) imply that $\hat{x}_1(t_0^+) = x(t_0)$. Similarly,

$$f_2(t_0^+) = -(x(t_0) - \hat{x}_2(t_0)) \tag{53}$$

and $\hat{x}_2(t_0^+) = x(t_0)$. We also note the fact that:

$$\dot{L}_R[f_1](t) = A_{\tilde{R}}L_R[f_1](t) + P^{-1}\tilde{R}g_1(t) \tag{54}$$

$$\begin{aligned}g_1(t) = \Phi_A(t, t_0)(\hat{x}_1(t_0^+) - f_1(t_0^+)) + \\ K^{t,t_0}[Bu_0 + Cv_0](t) + f_1(t)\end{aligned} \tag{55}$$

$$g_1(t_0^+) = x(t_0). \tag{56}$$

Similarly,

$$\dot{L}_S[f_2](t) = A_{\tilde{S}}L_S[f_2](t) + P^{-1}\tilde{S}g_2(t) \tag{57}$$

$$\begin{aligned}g_2(t) = \Phi_A(t, t_0)(\hat{x}_2(t_0^+) + f_2(t_0^+)) + \\ K^{t,t_0}[Bu_0 + Cv_0](t) - f_2(t)\end{aligned} \tag{58}$$

$$g_2(t_0^+) = x(t_0). \tag{59}$$

The conditions (52) and (53) ensure (50) and (51) hold true at $t = t_0^+$. To ensure (50) and (51) hold for all $t \in (t_0, t_1)$, we differentiate them and make the derivatives equal to zero for $t \in (t_0, t_1)$. Therefore differentiating (50) w.r.t. t we obtain:

$$\dot{f}_1 = Af_1 + P^{-1}\tilde{S}\zeta \tag{60}$$
$$\zeta = g_2 + L_S(f_2) \tag{61}$$
$$\dot{\zeta} = A_S\zeta + Bu_0 + Cv_0 - \dot{f}_2 + Af_2 \tag{62}$$
$$\zeta(t_0^+) = x(t_0). \tag{63}$$

At the same time, from (51) we obtain:

$$\dot{f}_2 = Af_2 + P^{-1}\tilde{R}\xi \tag{64}$$
$$\xi = g_1 - L_R(f_1) \tag{65}$$
$$\dot{\xi} = A_R\xi + Bu_0 + Cv_0 + \dot{f}_1 - Af_1 \tag{66}$$
$$\xi(t_0^+) = x(t_0). \tag{67}$$

From (22) and (66) we can conclude $\hat{x}_1 \equiv \xi$ for all $t \in (t_0, t_1)$ and similarly, $\hat{x}_2 \equiv \zeta$. Therefore, the coupled equations (60)-(67) can be rewritten as:

$$\dot{f}_1 = Af_1 + P^{-1}\tilde{S}\zeta \tag{68}$$
$$\dot{\zeta} = A_S\zeta + Bu_0 + Cv_0 - P^{-1}\tilde{R}\hat{x}_1 \tag{69}$$
$$\zeta(t_0^+) = x(t_0) \tag{70}$$
$$f_1(t_0^+) = x(t_0) - \hat{x}_1(t_0) \tag{71}$$

and,

$$\dot{f}_2 = Af_2 + P^{-1}\tilde{R}\xi \tag{72}$$
$$\dot{\xi} = A_R\xi + Bu_0 + Cv_0 + P^{-1}\tilde{S}\hat{x}_2 \tag{73}$$
$$\xi(t_0^+) = x(t_0) \tag{74}$$
$$f_2(t_0^+) = \hat{x}_2(t_0) - x(t_0). \tag{75}$$

Thus, the optimal f_1^* and f_2^* must satisfy the necessary conditions (68)-(75). This proves the second part of Theorem 3.6.

To prove that f_1^* and f_2^* satisfying (68)-(75) are a saddle point (hence Nash Equilibrium) for $\mathcal{J}$, we need to calculate the second order Gateaux differential of $\mathcal{J}$. In a similar fashion of deriving the first order Gateaux differential, one can find the second order Gateaux differential using (43).We do not present the details of this derivation due to space limitation, but one can verify that:

$$\frac{1}{2}\delta^2\mathcal{J}[f_1^*, f_2^*](h1, h2) = \mathbf{E}[J_1 - J_2 + J_3] \tag{76}$$

where

$$J_1 = \int_{t_0}^{t_1} (\|h_1\|_{\tilde{R}}^2 - \|\mathcal{G}_1[h_1]\|_{\tilde{S}}^2)dt \tag{77}$$

$$\mathcal{G}_1[h_1](t) = K_R^{t,t_0}[P^{-1}\tilde{R}h_1](t) - \Delta_R\Phi(t, t_0)h_1(t_0)$$

$$J_2 = \int_{t_0}^{t_1} (\|h_2\|_{\tilde{S}}^2 - \|\mathcal{G}_2[h_2]\|_{\tilde{R}}^2)dt \tag{78}$$

$$\mathcal{G}_2[h_2](t) = K_S^{t,t_0}[P^{-1}\tilde{S}h_2](t) - \Delta_S\Phi(t, t_0)h_2(t_0) \text{ and}$$

$$J_3 = 2\int_{t_0}^{t_1} \Big(\langle \mathcal{G}_2[h_2], h_1\rangle_{\tilde{R}} + \langle \mathcal{G}_1[h_1], h_2\rangle_{\tilde{S}}\Big)dt. \tag{79}$$

To prove that $\delta^2\mathcal{J}[f_1^*, f_2^*]$ is indefinite (i.e. depending on the direction (h_1, h_2), $\delta^2\mathcal{J}$ can be positive as well as negative), let us consider h_2^* to be of the form $h_2^*(t) = \Phi_A(t, t_0)a$ and $h_1^*(t) = \Phi(t, t_0)b$ for some constant vectors a and b. These choices imply $\mathcal{G}_2[h_2^*] \equiv 0$ and $\mathcal{G}_1[h_1^*] \equiv 0$. Therefore, $\delta^2\mathcal{J}[f_1^*, f_2^*](h_1^*, 0) = \int_{t_0}^{t_1} \|h_1^*\|_{\tilde{R}}^2 dt > 0$ for some a. Also, we have $\delta^2\mathcal{J}[f_1^*, f_2^*](0, h_2^*) = -\int_{t_0}^{t_1} \|h_2^*\|_{\tilde{S}}^2 dt < 0$ for some b. This proves that (f_1^*, f_2^*) is a saddle point of $\mathcal{J}$.

The uniqueness of (f_1^*, f_2^*) is a direct consequence of the fact that they satisfy certain linear differential equations with fixed boundary conditions.

REFERENCES

[1] T. Basar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1995, vol. 200.

[2] M. R. James and J. Baras, "Partially observed differential games, infinite-dimensional hamilton-jacobi-isaacs equations, and nonlinear h_ control," *SIAM Journal on Control and Optimization*, vol. 34, no. 4, pp. 1342–1364, 1996.

[3] T. Başar and P. Bernhard, *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media, 2008.

[4] J. Engwerda, *LQ dynamic optimization and differential games*. John Wiley & Sons, 2005.

[5] W. H. Fleming and D. Hernández-Hernández, "On the value of stochastic differential games," *Commun. Stoch. Anal*, vol. 5, no. 2, pp. 341–351, 2011.

[6] R. Isaacs, *Differential games: a mathematical theory with applications to warfare and pursuit, control and optimization*. Courier Corporation, 1999.

[7] A. W. Starr and Y.-C. Ho, "Nonzero-sum differential games," *Journal of Optimization Theory and Applications*, vol. 3, no. 3, pp. 184–206, 1969.

[8] ——, "Further properties of nonzero-sum differential games," *Journal of Optimization Theory and Applications*, vol. 3, no. 4, pp. 207–219, 1969.

[9] M. Foley and W. Schmitendorf, "On a class of nonzero-sum, linear-quadratic differential games," *Journal of Optimization Theory and Applications*, vol. 7, no. 5, pp. 357–377, 1971.

[10] A. Weeren, J. Schumacher, and J. Engwerda, "Asymptotic analysis of linear feedback Nash equilibria in nonzero-sum linear-quadratic differential games," *Journal of Optimization Theory and Applications*, vol. 101, no. 3, pp. 693–722, 1999.

[11] T. Basar, "On the uniqueness of the Nash solution in linear-quadratic differential games," *International Journal of Game Theory*, vol. 5, no. 2-3, pp. 65–90, 1976.

[12] J. Cruz Jr and C. Chen, "Series Nash solution of two-person, nonzero-sum, linear-quadratic differential games," *Journal of Optimization Theory and Applications*, vol. 7, no. 4, pp. 240–257, 1971.

[13] D. H. Jacobson, "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games," *Automatic Control, IEEE Transactions on*, vol. 18, no. 2, pp. 124–131, 1973.

[14] P. Bernhard, "Linear-quadratic, two-person, zero-sum differential games: necessary and sufficient conditions," *Journal of Optimization Theory and Applications*, vol. 27, no. 1, pp. 51–69, 1979.

[15] M. Simaan and J. B. Cruz Jr, "On the Stackelberg strategy in nonzero-sum games," *Journal of Optimization Theory and Applications*, vol. 11, no. 5, pp. 533–555, 1973.

[16] T. Basar, "A counterexample in linear-quadratic games: Existence of nonlinear Nash solutions," *Journal of Optimization Theory and Applications*, vol. 14, no. 4, pp. 425–430, 1974.

[17] T. Basar *et al.*, "Lecture notes on non-cooperative game theory," *Game Theory Module of the Graduate Program in Network Mathematics*, pp. 3–6, 2010.

[18] T. E. Duncan and B. Pasik-Duncan, "Some stochastic differential games with state dependent noise," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 3773–3777.

[19] D. Maity and J. S. Baras, "Strategies for two-player differential games with costly information," in *2016 13th International Workshop on Discrete Event Systems (WODES)*. IEEE, 2016, pp. 211–216.