

Linear Quadratic Games with Costly Measurements

Dipankar Maity¹, Achilleas Anastasopoulos², and John S. Baras¹

Abstract— In this work we consider a stochastic linear quadratic two-player game. The state measurements are observed through a switched noiseless communication link. Each player incurs a finite cost every time the link is established to get measurements. Along with the usual control action, each player is equipped with a switching action to control the communication link. The measurements help to improve the estimate and hence reduce the quadratic cost, but at the same time the cost is increased due to switching. We study the subgame perfect equilibrium control and switching strategies for the players. We show that the problem can be solved in a two-step process by solving two dynamic programming problems. The first step corresponds to solving a dynamic programming for the control strategy and the second step involves solving another dynamic programming for the switching strategy.

I. INTRODUCTION

Linear quadratic (LQ) stochastic games have attracted a great deal of attention in the control and related community due to its wide applicability in stochastic control, minimax control, multi-agent systems and economics [1], [2], [3], [4], [5], [6], [7]. There is a well established notion of (Nash) equilibrium (NE) strategies for static games, and in dynamic games there are refinements of NE known as subgame perfect equilibria (SPE). Closed form solutions for these NE (or SPE) may generally not exist or hard to compute if one such exists. Among the various classes of dynamic games, LQ games exhibit a closed form expression for SPE, and it is characterized by some Riccati equations. Necessary and sufficient conditions for the NE strategies of LQ games are studied in [3], [8], [5]. Contrary to the prior belief, [9] shows existence of nonlinear control strategies for LQ games.

Amongst the vast majority of the prior works, the underlying assumption is the availability of free observations. Dynamic games are studied with either open-loop strategy (i.e. only measurement is the initial state) or feedback strategies where the observation is available freely at any time. Challenges emerge when the measurements are on demand, but costly. This adds an extra layer of decision making, for the players, because now they have to control the system as well as ask for measurements.

In this work, we consider a class of two-player linear quadratic stochastic games of finite horizon. The game dynamics are partially observable. Contrary to the existing

¹The authors are with the Department of Electrical and Computer Engineering and The Institute for Systems Research, University of Maryland, College Park, MD, USA. Email: dmaity@umd.edu, baras@umd.edu. Research partially supported by DARPA grant W911NF-14-1-0384 through ARO, NSF grant CNS-1544787, and DARPA STTR contract with Boston Engineering Corporation.

² The author is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA. Email: anastas@umich.edu

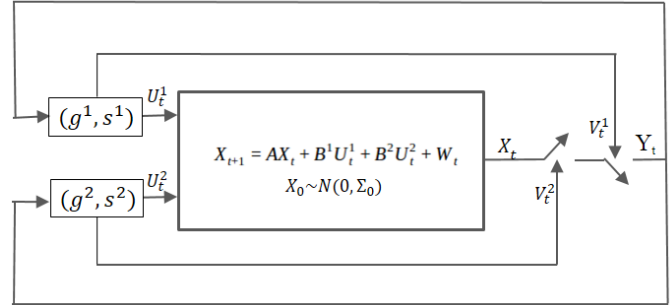


Fig. 1. Schematic of the system. Each player has to select their controller strategy g^i and switching strategy s^i . All the links are noiseless and delay-free.

literature, the observations are not freely available. Each observation requires a finite cost for establishing a link for communication. This communication link is assumed to be noiseless but operated by two switches (Figure 1), one for each player. The link is established only when both the players are willing for it, and they both get the actual state measurement at that time. Consequently, there is an apparent trade off between cost of obtaining state measurement and the estimation quality.

In this game, the players can make a precise estimate of the state if they establish the link at every time instance. However, since the link establishment is costly, they can compromise the estimation accuracy in exchange of the cost for accessing the measurement. Therefore, the problem is to optimally decide when to establish the link and how to use the acquired measurement in order to minimize their individual costs. Since, in general, the players will have different preferences over the time instances when they want to acquire the measurement, they have to come to an agreement when to actually establish the link.

The closest work on the similar game framework has been studied in [10] where the authors studied zero-sum stochastic differential LQ games. However, the selection for switching times were performed in an collaborative way rather than being the outcome of a strategic interaction. The major digression of our work from [10] is that we consider an explicit game for the switching strategy as well. We express the switch as a Boolean control action and seek for SPE for both control and switching strategies.

Our contributions are as follows:

- (a) We study the SPE of this dynamic game and show that they can be found through a two-step process. Specifically, in the first step we fix the switching strategy and study

the SPE for control strategies. The study shows that the control strategy is linear in estimated state, where the gain is characterized with two backward Riccati equations which can be computed offline. Moreover, the Riccati equations do not depend on the switching strategy.

(b) Regarding the equilibrium switching strategy, we provide a backward recursive algorithm to find all SPE where value functions need only be computed over a finite and quadratically-sized (in the duration of the game) set.

(c) Further regarding the equilibrium switching strategy, we show that there are many equilibria among which there is one that is strictly preferable by both users and has a Markovian structure. It is found in our study that a strictly preferable switching strategy for a player not only depends on their own cost-to-go, but also depends on the cost-to-go for the opponent.

The remaining of the paper is organized as follows: The problem formulation is provided in Section II, Section III contains the results on the SPE of the control strategy, SPE for the switching strategy and its offline computation are analyzed in Section IV. Finally we conclude our work in Section VI.

II. PROBLEM FORMULATION

In the discrete time Gauss-Markov setting, we consider the following linear dynamics of the state X_t :

$$X_{t+1} = AX_t + B^1U_t^1 + B^2U_t^2 + W_t \quad (1)$$

where $X_t \in \mathcal{X} = \mathbb{R}^n$, and $U_t^i \in \mathcal{U}^i = \mathbb{R}^{m_i}$ denotes the action of player i . $W_t \in \mathbb{R}^n$ is a Gaussian noise with $\mathbb{E}[W_t] = 0$ and $\mathbb{E}[W_t W_s^T] = S\delta_{t-s}$ (δ_t is the Kronecker delta.), and $X_0 \sim \mathcal{N}(0, \Sigma_0)$.

There are two additional actions (switching actions) $V_t^1 \in \{0, 1\}$ and $V_t^2 \in \{0, 1\}$. These switching actions control a switch (switch closes if both are equal to 1) and the observation available to both users is $Y_t \in \mathcal{X} \cup \{e\}$ with

$$Y_t = \begin{cases} X_t & , V_t^1 = V_t^2 = 1 \\ e & , \text{else,} \end{cases} \quad (2)$$

where “e” denotes an erasure. The evolution of random variables in period t is assumed to be $\dots X_t \rightarrow (V_t^1, V_t^2) \rightarrow Y_t \rightarrow (U_t^1, U_t^2) \dots$

The information available at time t to player i before she takes the switching action V_t^i is

$$I_t = (Y^{t-1}, U^{1,t-1}, U^{2,t-1}, V^{1,t-1}, V^{2,t-1}), \quad (3)$$

and the information available at time t to player i before she takes the control action U_t^i is

$$\bar{I}_t = (I_t, V_t^1, V_t^2, Y_t). \quad (4)$$

As a result, the actions have the functional form

$$V_t^i = s_t^i(I_t), \quad i = 1, 2, \quad (5a)$$

$$U_t^i = g_t^i(\bar{I}_t), \quad i = 1, 2, \quad (5b)$$

where by $g^i = (g_t^i)_{t=0}^{T-1}$, $s^i = (s_t^i)_{t=0}^{T-1}$, we denote the control and switching strategies of player i . $\forall t \in \{0, \dots, T-1\}$, let us denote an I_t measurable random variable $\Delta_t = V_t^1 \cdot V_t^2$.

The individual cost that each player needs to minimize is quadratic in state and action, and it also depends on the switching actions V_t^1 and V_t^2 . We consider a game for a finite duration ($\{0, \dots, T\}$) and the per-stage costs are explicitly written as:

$$C_t^i(x_t, u_t^1, u_t^2, v_t^1, v_t^2) = \|x_t\|_{Q^i}^2 + \|u_t^i\|_{Q^{ii}}^2 + \|u_t^j\|_{Q^{ij}}^2 + \lambda_i(v_t^i \cdot v_t^j) \quad (6)$$

for all $t \in \{0, \dots, T-1\}$ and

$$C_T^i(x_T, u_T^1, u_T^2, v_T^1, v_T^2) = \|x_T\|_{Q^i}^2. \quad (7)$$

The quantity $\lambda_i > 0$ is the cost paid by player i when both the players attempt to close the switch and they observe the state information X_t . Therefore the average cost over the time horizon $\{0, \dots, T\}$ is represented as,

$$J^i(\sigma^1, \sigma^2) = \sum_{t=0}^T \mathbb{E}[C_t^i(X_t, U_t^1, U_t^2, V_t^1, V_t^2)] \quad (8)$$

where $\sigma^i = (g^i, s^i)$ denotes the strategy of the player i that corresponds to control strategy g^i and switching strategy s^i .

The objective of player i is:

$$\min_{\sigma^i} J^i(\sigma^1, \sigma^2) = \min_{s^i} \{ \min_{g^i} J^i(\sigma^1, \sigma^2) \} \quad (9)$$

III. SUBGAME PERFECT CONTROL STRATEGY

For dynamic games with complete information the appropriate equilibrium concept is a refinement of NE called the subgame perfect equilibrium (SPE). A strategy profile (σ^1, σ^2) is a SPE if the restriction of (σ^1, σ^2) to any proper subgame of the original game constitutes a NE [11, pp. 94].

We seek to characterize the SPE $(\sigma^{1*}, \sigma^{2*})$ for this switched LQG game. Moreover, we will show that among the multiple SPE, there exists one that simultaneously minimizes the cost for both users among all SPE and thus it will be the preferable SPE solution of this game. In this section, we study the SPE control strategy for both the players.

Theorem 3.1: For any switching profile (s^1, s^2) of the players, the SPE control strategy g^{i} has the following structure:*

$$U_t^i = g_t^{i*}(\bar{I}_t) = -L_t^i \hat{X}_t, \quad (10)$$

where

$$\hat{X}_t = \begin{cases} A\hat{X}_{t-1} + B^1U_{t-1}^1 + B^2U_{t-1}^2, & V_t^1 \cdot V_t^2 = 0 \\ X_t, & V_t^1 \cdot V_t^2 = 1. \end{cases} \quad (11)$$

Furthermore, the cost-to-go incurred by player i under the SPE control strategy at any time step k is given by,

$$\mathbb{J}_k^{i*}(\bar{I}_k) = \|\hat{X}_k\|_{P_k^i}^2 + \mathbb{E} \left[\|E_T\|_{Q^i}^2 + \sum_{t=k}^{T-1} (\|E_t\|_{Q^i}^2 + \lambda_i \Delta_t) + \sum_{t=k}^{T-2} \Delta_{t+1} \|AE_t + W_t\|_{P_{t+1}^i}^2 \mid \bar{I}_k \right] \quad (12)$$

where $E_t = X_t - \hat{X}_t$. The matrices L_t^i and P_t^i depend only on the game parameters A, B^i, Q^i, Q^{ii} and Q^{ij} (see [12])

and thus, can be calculated offline without the knowledge of the switching strategy profile.

Proof: The proof of this theorem is provided in the unabridged version of this work [12]. ■

To maintain brevity $\mathbb{J}_k^{i*}(\bar{I}_k)$ will be denoted as \mathbb{J}_k^{i*} . From this point onward we will set $\Delta_T = 0$ and write (12) in compact form as

$$\mathbb{J}_k^{i*}(\bar{I}_k) = \mathbb{E} \left[\sum_{t=k}^{T-1} (\|E_t\|_{Q^i}^2 + \Delta_{t+1} \|AE_t + W_t\|_{P_{t+1}^i}^2 + \lambda_i \Delta_t) + \|E_T\|_{Q^i}^2 | \bar{I}_k \right] + \|\hat{X}_k\|_{P_k^i}^2 \quad (13)$$

It should be noted that in Theorem 3.1, the g^{i*} depends on the given switching strategy (s^1, s^2) through the \hat{X}_t .

There are several remarks to be made at this point. The stochastic control version of the same problem (i.e. single player single objective) is a modified Kalman filtering problem where the observations are available on demand after paying certain cost λ per observation. Therefore, the decision of switching will solely depend on the influence of switching on the error covariance matrix. This is a side result of our work and is not provided due to space limitation.

From Theorem 3.1, $\min_{\sigma^i} J^i(\sigma^1, \sigma^2) = \min_{s^i} \mathbb{E}[\mathbb{J}_0^{i*}]$. Therefore, the total cost incurred by player i with control strategy profile (g^{1*}, g^{2*}) is $\mathbb{E}[\mathbb{J}_0^{i*}]$. Hence, the total cost incurred with the switching is:

$$\mathbb{E}[\mathbb{J}_0^{i*}] = \mathbb{E}[\|\hat{X}_0\|_{P_0^i}^2] + \mathbb{E} \left[\sum_{t=0}^{T-1} (\|E_t\|_{Q^i}^2 + \Delta_{t+1} \|AE_t + W_t\|_{P_{t+1}^i}^2 + \lambda_i \Delta_t) + \|E_T\|_{Q^i}^2 \right]. \quad (14)$$

Another remark that is apparent here is that the SPE control strategy is completely characterized by the pair of matrices (P_t^1, P_t^2) which is uniquely determined by backward dynamic equations, and independent of switching strategy.

IV. SUBGAME PERFECT SWITCHING STRATEGY

In this section we complete the procedure for finding the SPE of this game by focusing on the switching strategies. We will do that by considering the backward induction process to find SPE and reduce the cost-to-go functions into a simpler and more tractable form (compared to the one in (13)).

In this problem the switching action is taken first at time k based on the knowledge I_k and then the augmented knowledge $\bar{I}_k = (I_k, V_k^1, V_k^2, Y_k)$ is used to select the control strategies U_k^i . In order to visualize it, one might break the time period $[k, k+1]$ into two halves where in the first half, switching action is performed and in the second half, control action is performed. In Theorem 3.1, \mathbb{J}_k^{i*} is the optimal cost-to-go after the switching decision has been taken at time k .

The actual (before switching action is taken) cost-to-go at stage k is:

$$\mathbb{V}_k^i(I_k) = \mathbb{E} \left[\sum_{t=k}^T C_t^i(X_t, U_t^1, U_t^2, V_t^1, V_t^2) | I_k \right] \quad (15)$$

and the optimization (game) variables are control U_t^i and switching V_t^i for all $t \geq k$.

Due to the fact $\bar{I}_t \supseteq I_t$ for all t , we can write

$$\begin{aligned} \mathbb{V}_k^i(I_k) &= \mathbb{E} \left[\mathbb{E} \left[\sum_{t=k}^T C_t^i(X_t, U_t^1, U_t^2, V_t^1, V_t^2) | \bar{I}_k \right] | I_k \right] \quad (16) \\ &= \mathbb{E}[\mathbb{J}_k^i(g^1, g^2) | I_k] \end{aligned}$$

where $\mathbb{J}_k^i(g^1, g^2) = \mathbb{E} \left[\sum_{t=k}^T C_t^i(X_t, U_t^1, U_t^2, V_t^1, V_t^2) | \bar{I}_k \right]$.

Since each player is interested in minimizing their cost, they are interested in $\min_{s^i, g^i} \mathbb{V}_k^i(I_k)$ at every stage k (finally they want to minimize $\mathbb{V}_0^i(I_0)$).

We can write,

$$\min_{s^i, g^i} \mathbb{V}_k^i(I_k) = \min_{s^i} \{ \min_{g^i} \mathbb{V}_k^i(I_k) \} = \min_{s^i} \mathbb{E}[\mathbb{J}_k^{i*} | I_k]. \quad (17)$$

We substitute the expression of \mathbb{J}_k^{i*} from Theorem 3.1 into (17), but before that, let us define,

$$M_t = \mathbb{E}[E_t E_t' | \bar{I}_t] = (1 - \Delta_t)(AM_{t-1}A' + S) \quad (18)$$

where $AM_{-1}A' + S = \Sigma_0$ (since $X_0 \sim \mathcal{N}(0, \Sigma_0)$).

We also define $M_{t|t-1} = AM_{t-1}A' + S$. Note that $M_{t|t-1}$ is I_t measurable whereas M_t is \bar{I}_t measurable. M_t and $M_{t|t-1}$ are related as follows:

$$M_t = (1 - \Delta_t)M_{t|t-1} \quad (19)$$

Now let us consider the k -th stage cost \mathbb{J}_k^{i*} .

$$\begin{aligned} \mathbb{J}_k^{i*} &= \mathbb{E} \left[\sum_{t=k}^{T-1} (\|E_t\|_{Q^i}^2 + \Delta_{t+1} \|AE_t + W_t\|_{P_{t+1}^i}^2 + \lambda_i \Delta_t) + \|E_T\|_{Q^i}^2 | \bar{I}_k \right] + \|\hat{X}_k\|_{P_k^i}^2 \\ &= \mathbb{E} \left[\sum_{t=k}^{T-1} (tr(Q^i M_t + \Delta_{t+1}(AM_t A' + S)P_{t+1}^i) + \lambda_i \Delta_t) + tr(Q^i M_T) | \bar{I}_k \right] + \|\hat{X}_k\|_{P_k^i}^2 \\ &= \mathbb{E} \left[\sum_{t=k}^{T-1} (tr((1 - \Delta_t)Q^i M_{t|t-1} + \Delta_{t+1}M_{t+1|t}P_{t+1}^i) + \lambda_i \Delta_t) + tr(Q^i M_T) | \bar{I}_k \right] + \|\hat{X}_k\|_{P_k^i}^2 \quad (20) \end{aligned}$$

Let us define $\mathcal{V}_k^i(I_k) = \mathbb{E}[\mathbb{J}_k^{i*} | I_k]$. Therefore,

$$\begin{aligned} \mathcal{V}_k^i(I_k) &= \mathbb{E} \left[\sum_{t=k}^{T-1} (tr((1 - \Delta_t)Q^i M_{t|t-1} + tr(\Delta_{t+1}M_{t+1|t}P_{t+1}^i) + \lambda_i \Delta_t) + tr(Q^i M_T) | I_k \right] + \mathbb{E}[\|\hat{X}_k\|_{P_k^i}^2 | I_k] \quad (21) \end{aligned}$$

One can verify that $\hat{X}_t = \mathbb{E}[X_t | \bar{I}_t]$ and we define the prediction variable as $\hat{X}_{t+1|t} = \mathbb{E}[X_{t+1} | I_t]$. Therefore,

$$\hat{X}_t = \hat{X}_{t|t-1} + \Delta_t(AE_{t-1} + W_{t-1}) \quad (22)$$

Using (22), we get

$$\begin{aligned} \mathcal{V}_k^i(I_k) &= \mathbb{E} \left[\sum_{t=k}^{T-1} (tr((1 - \Delta_t)Q^i M_{t|t-1} + tr(\Delta_t M_{t|t-1} P_t^i) + \lambda_i \Delta_t) + tr(Q^i M_T) | I_k \right] + \|\hat{X}_{k|k-1}\|_{P_k^i}^2 \quad (23) \end{aligned}$$

The selection of switching strategy $s_k^i(I_k)$ has no effect of $\hat{X}_{k|k-1}$ and hence it does not play any role in the game at stage k .

Let us define an instantaneous cost:

$$\bar{C}_t^i(M_{t|t-1}, \Delta_t) = (1 - \Delta_t)tr(Q^i M_{t|t-1}) + \Delta_t tr(M_{t|t-1} P_t^i) + \lambda_i \Delta_t. \quad (24)$$

With slight abuse of notation, after neglecting the $\hat{X}_{k|k-1}$ term, we obtain,

$$\mathcal{V}_k^i(I_k) = \mathbb{E} \left[\sum_{t=k}^T \bar{C}_t^i(M_{t|t-1}, \Delta_t) \mid I_k \right]. \quad (25)$$

Therefore,

$$\min_{s^i, g^i} \mathbb{V}_k^i(I_k) = \min_{s^i} \mathcal{V}_k^i(I_k) = \min_{s^i} \mathbb{E} \left[\sum_{t=k}^T \bar{C}_t^i(M_{t|t-1}, \Delta_t) \mid I_k \right].$$

Let us denote:

$$\mathcal{V}_k^{i*} = \min_{s^i} \mathcal{V}_k^i(I_k). \quad (26)$$

Let us perform the similar backward induction to find the SPE for the switching strategies. Note at time T , there is no action to optimize and

$$\mathcal{V}_T^i(I_T) = \mathbb{E} \left[\bar{C}_T^i(M_{T|T-1}, \Delta_T) \mid I_T \right] = \mathbb{E} \left[tr(Q^i M_T) \mid I_T \right].$$

Let us define

$$\mathcal{V}_T^{i*} = \mathcal{V}_T^i(I_T) = \mathbb{E} \left[tr(Q^i M_T) \mid I_T \right]. \quad (27)$$

Similarly, at $T-1$,

$$\mathcal{V}_{T-1}^i(I_{T-1}) = \mathbb{E} \left[\bar{C}_{T-1}^i(M_{T-1|T-2}, \Delta_{T-1}) + tr(Q^i M_T) \mid I_{T-1} \right]. \quad (28)$$

Using $M_{T-1} = (1 - \Delta_{T-1})M_{T-1|T-2} + \Delta_{T-1}M_T$ and $M_T = AM_{T-1}A' + S$,

$$\mathcal{V}_{T-1}^i(I_{T-1}) = \mathbb{E} \left[\bar{C}_{T-1}^i(M_{T-1|T-2}, \Delta_{T-1}) + (1 - \Delta_{T-1})tr(Q^i (AM_{T-1|T-2}A')) + tr(Q^i S) \mid I_{T-1} \right]$$

If (s_{T-1}^1, s_{T-1}^2) is a SPE strategy at time $T-1$ then

$$\mathcal{V}_{T-1}^i(I_{T-1}) \Big|_{(s_{T-1}^i, s_{T-1}^j)} \leq \mathcal{V}_{T-1}^i(I_{T-1}) \Big|_{(s_{T-1}^i, s_{T-1}^j)} \quad (29)$$

$\forall s_{T-1}^i$ and for both $i = 1, 2; j = 1, 2$ and $i \neq j$.

Using the above definition of SPE, $s_{T-1}^i(I_{T-1}) = 0$ for $i = 1, 2$ is an equilibrium strategy since unilateral change from 0 to 1 does not change the cost for any player. However, there might be other equilibria (in this case only $(1, 1)$) which produces lower cost for the above cost function.

It is straightforward to show that the equilibrium strategy at $T-1$ is

$$s_{T-1}^{i*}(I_{T-1}) = \begin{cases} 1 & \text{if } \alpha_{T-1}^i \leq tr(Q^i (AM_{T-1|T-2}A')) \\ 0 & \text{otherwise} \end{cases} \quad (30)$$

where $\alpha_{T-1}^i = \bar{C}_{T-1}^i(M_{T-1|T-2}, 1) - \bar{C}_{T-1}^i(M_{T-1|T-2}, 0)$

From (30) we notice that $(1, 0)$ and $(0, 1)$ can also be an equilibrium strategy. However those equilibria are equivalent to $(0, 0)$ in the sense that they produce the same cost-to-go \mathcal{V}_{T-1}^{i*} for both $i = 1, 2$. Therefore, we will restrict our attention on two equilibria $(0, 0)$ and $(1, 1)$

As a remark, it is pointed out that adding an infinitesimal switching cost ϵ_i for every time player i requests for a switching (irrespective of whether the switch was closed or not) will ensure that $(0, 1)$ and $(1, 0)$ is never an SPE.

Let us notice that when $\alpha_{T-1}^i = tr(Q^i (AM_{T-1|T-2}A'))$, then $s_{T-1}^{i*} = 0$ or 1 , both produces the same cost-to-go value. Under such situations, all possible switching actions are equivalent. In order to obliterate such instances we make the following assumption:

Assumption 4.1: If $\alpha_{T-1}^i = tr(Q^i (AM_{T-1|T-2}A'))$, $s_{T-1}^{i*}(I_{T-1}) = 0$ for all possible history I_{T-1} . Therefore, (30) is modified as follows:

$$s_{T-1}^{i*}(I_{T-1}) = \begin{cases} 1 & \text{if } \alpha_{T-1}^i \leq tr(Q^i (AM_{T-1|T-2}A')) \\ 0 & \text{otherwise} \end{cases} \quad (31)$$

Irrespective of whether SPE $s_{T-1}^{i*}(I_{T-1})$ is 0 or 1, the optimal cost-to-go \mathcal{V}_{T-1}^{i*} depends only on M_{T-2} and also the best SPE strategy (that produces the least cost among all SPE) $s_{T-1}^{i*}(I_{T-1})$ depends only on M_{T-2} (or $M_{T-1|T-2}$).

Therefore, we hypothesize the following:

Claim 4.2: For any k , there exists a $s_k^{i*}(I_k)$ that depends only on M_{k-1} and produces the least cost-to-go among all SPE. Hence $\mathcal{V}_k^{i*} \equiv \mathcal{V}_k^{i*}(M_{k-1})$ (i.e. \mathcal{V}_k^{i*} only depends on M_{k-1}).

Proof: The hypothesis is true for $k = T, T-1$. Let us assume it is true for some $k+1 \leq T$, i.e. $\mathcal{V}_{k+1}^{i*} \equiv \mathcal{V}_{k+1}^{i*}(M_k)$. Therefore,

$$\mathcal{V}_k^{i*} = \min_{s^i} \sum_{t=k}^T \mathbb{E} \left[\bar{C}_t^i(M_{t|t-1}, \Delta_t) \mid I_k \right]$$

Then using a dynamic programming argument,

$$\begin{aligned} \mathcal{V}_k^{i*} &= \min_{s_k^i} \mathbb{E} \left[\bar{C}_k^i(M_{k|k-1}, \Delta_k) + \mathcal{V}_{k+1}^{i*}(M_k) \mid I_k \right] \\ &= \min_{s_k^i} \mathbb{E} \left[\bar{C}_k^i(M_{k|k-1}, \Delta_k) + \mathcal{V}_{k+1}^{i*}((1 - \Delta_k)M_{k|k-1}) \mid I_k \right] \end{aligned} \quad (32)$$

From (32), the best equilibrium strategy $s_k^{i*}(I_k) = 1$ if

$\bar{C}_k^i(M_{k|k-1}, 1) + \mathcal{V}_{k+1}^{i*}(\mathbf{0}) < \bar{C}_k^i(M_{k|k-1}, 0) + \mathcal{V}_{k+1}^{i*}(M_{k|k-1})$ (similar to assumption 4.1, we only consider the strict inequality), otherwise $s_k^i(I_k) = 0$.

Therefore $s_k^{i*}(I_k)$ requires only the knowledge of M_{k-1} and hence from (32), $\mathcal{V}_k^i(I_k) \equiv \mathcal{V}_k^{i*}(M_{k-1})$ ■

For this class of games, there always exists a Markovian SPE switching strategy and a Markovian SPE control strategy which produce the least cost-to-go among all SPE. Though, there might be other non-Markovian SPE strategies which produce the same cost, however, due to the claim 4.2, it is sufficient to consider only the Markovian strategies to find the best SPE corresponding to the least cost-to-go.

A. Offline Calculation of $\mathcal{V}_k^{i*}(M_{k-1})$

In the following we define how the players can take the decision online by using some stored offline functions (value functions).

Let us define $\mathcal{V}_k^{i*}(M)$ in the following manner:

$$\mathcal{V}_T^{i*}(M) = \bar{C}_T^i(M, 0). \quad \forall M \text{ and } i = 1, 2 \quad (33)$$

and

$$\mathcal{V}_k^{i*}(M) = \begin{cases} \bar{C}_k^i(M, 1) + \mathcal{V}_{k+1}^{i*}(\mathbf{0}) & \text{if } \vartheta(k, M) > 1, \\ \bar{C}_k^i(M, 0) + \mathcal{V}_{k+1}^{i*}(AMA' + S) & \text{otherwise.} \end{cases} \quad (34)$$

where $\vartheta(k, M) = \min\{\vartheta^1(k, M), \vartheta^2(k, M)\}$, and

$$\vartheta^i(k, M) = \frac{\bar{C}_k^i(M, 0) + \mathcal{V}_{k+1}^{i*}(AMA' + S)}{\bar{C}_k^i(M, 1) + \mathcal{V}_{k+1}^{i*}(\mathbf{0})} \quad (35)$$

By construction, if $\mathcal{V}_{k+1}^{i*}(\cdot)$ denotes the minimum cost-to-go (for the subgame starting at $k+1$) among the SPE, $\mathcal{V}_k^{i*}(\cdot)$ defined in (34) provides the minimum cost-to-go at stage k for player i . Therefore, by backward inductions, $\mathcal{V}^{i*}(\cdot)$ denotes the cost-to-go function along an SPE that simultaneously minimizes the cost-to-go for both players.

Claim 4.3: For any k, M and history I_k , the best switching strategy (SPE) is given by $s_k^{i*}(I_k) = 1$ for $i = 1, 2$ if and only if,

$$\bar{C}_k^i(M, 1) + \mathcal{V}_{k+1}^{i*}(\mathbf{0}) < \bar{C}_k^i(M, 0) + \mathcal{V}_{k+1}^{i*}(AMA' + S). \quad (36)$$

Otherwise $s_k^{i*}(I_k) = 0$ for $i = 1, 2$.

Proof: \Rightarrow is trivially true.

\Leftarrow : First, notice that we have established $s_k^i(I_k) = 0$ is an SPE strategy for all k, I_k . Now let us assume that at some k, M , (36) holds, then if player i selects a strategy such that $s_k^i(I_k) = 0$, then the cost-to-go for player i with any strategy profile $((s^1)_{k+1}^T, (s^2)_{k+1}^T)$ from time $k+1$ onward is

$$\begin{aligned} & \bar{C}_k^i(M, 0) + \mathcal{V}_{k+1}^i(AMA' + S, (s^1)_{k+1}^T, (s^2)_{k+1}^T) \\ & \geq \bar{C}_k^i(M, 0) + \mathcal{V}_{k+1}^{i*}(AMA' + S) \\ & > \bar{C}_k^i(M, 1) + \mathcal{V}_{k+1}^{i*}(\mathbf{0}) \end{aligned} \quad (37)$$

Therefore, unilateral deviation is harmful (strictly non-profitable) for the player i , and that allows us to conclude $s_k^i(I_k) = 1$ for $i = 1, 2$ is an equilibrium for (k, M) . Therefore $(s_k^1(I_k), s_k^2(I_k)) = (0, 0)$ and $(1, 1)$ both are equilibria. However, the cost-to-go by selecting $(1, 1)$ is strictly lesser than selecting $(0, 0)$, and this is, therefore, preferable by the players. ■

Note that, (34) can be calculated and stored offline and (36) can be evaluated online using the stored values.

Equation (36) is equivalent to:

$$\lambda_i < \mathcal{V}_{k+1}^{i*}(AMA' + S) - \mathcal{V}_{k+1}^{i*}(\mathbf{0}) - \text{tr}((P_k^i - Q^i)(AMA' + S)) \quad (38)$$

which shows a threshold policy for SPE switching.

We note that $M_{0|-1} = \Sigma_0$, therefore at time 0 we only need the value $\mathcal{V}_0^{i*}(\Sigma_0)$ not the function $\mathcal{V}_0^{i*}(\cdot)$ in the entire space of symmetric positive semidefinite matrices. In order to

decide $(s_0^{1*}(I_0), s_0^{2*}(I_0))$ we need to know only four values $\mathcal{V}_1^{i*}(\mathbf{0}), \mathcal{V}_1^{i*}(A\Sigma_0A' + S)$ for $i = 1, 2$. Therefore, given the variance of X_0 , we need to store only finite number of values to characterize all the value functions for a finite duration game.

Claim 4.4: The maximum number of values (value function evaluations) needed to be stored to calculate the switching strategies for entire game of duration $[0, T]$ is $T(T+3)$.

Proof: Let at stage k , $M_{k|k-1}$ (or M_{k-1}) takes n_k number of possible distinct values based on all possible previous history I_k . Therefore to determine the switching at time k , we need to make n_k comparison tests (36) and for each test the $\mathcal{V}_{k+1}^{i*}(\mathbf{0})$ term is common. Therefore we need to evaluate the value function only at $n_k + 1$ number of points at time k .

For the switching pair $(1, 1)$, $M_k = 0$ (or $M_{k+1|k} = S$) and for any other possible switching profile at stage k , $M_k = M_{k|k-1}$. Therefore at stage $k+1$, n_{k+1} will be at most $n_k + 1$ (and n_{k+1} possible values of M_k .) Therefore,

$$n_{k+1} \leq n_k + 1 \quad (39)$$

with $n_0 = 1$, we get $n_k \leq k + 1$.

Total value function evaluations to be stored =

$$2 * \sum_{k=0}^{T-1} (n_k + 1) \leq T(T+3).$$

The factor 2 in above equation is due to the fact that we have to evaluate the value functions for both the players. ■

Remark 4.5: A switching is performed only when it strictly reduces the cost-to-go for both users. Therefore, each switching minimizes the welfare cost-to-go. However, the converse is not necessarily true i.e. a switching with a potential to reduce the welfare cost-to-go may not always be performed.

V. SIMULATION RESULTS

We consider the following two-dimensional system to illustrate our analysis that has been carried out in the preceding sections.

$$X_{k+1} = \begin{bmatrix} 0.4 & 0.8 \\ -0.8 & 1 \end{bmatrix} X_k + U_k^1 - U_k^2 + W_k$$

where $X_k, U_k^1, U_k^2, W_k \in \mathbb{R}^2$ for all k . $W_k \sim \mathcal{N}(0, 0.25\mathbf{I})$. The observation cost parameters $\lambda_1 = 1$ and $\lambda_2 = 1.5$.

For the cost (6), the following parameters are taken:

$$Q^1 = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.7 \end{bmatrix}, Q^2 = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.2 \end{bmatrix},$$

$$Q^{12} = Q^{21} = \mathbf{0} \text{ and } Q^{11} = Q^{22} = \mathbf{I}.$$

One can show that $L_{t-1}^i = (-1)^{i-1} P_t^i (\mathbf{I} + P_t^1 + P_t^2)^{-1} A$. By denoting $P_t = \mathbf{I} + P_t^1 + P_t^2$, one can verify:

$$\begin{aligned} P_t^i &= Q^i + A' P_{t+1}^{-1} P_{t+1}^i (P_{t+1}^i + 1) P_{t+1}^{-1} A \\ P_T^i &= Q^i \end{aligned}$$

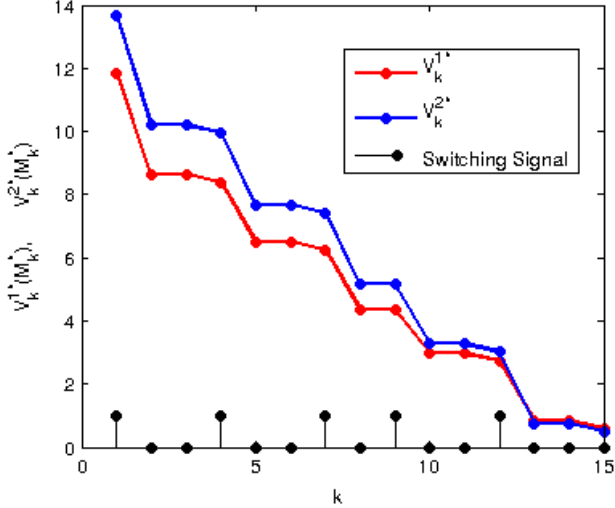


Fig. 2. The red and blue lines plot $V_k^{i*}(M^*)$ w.r.t k for $i = 1$ and 2 respectively. M^* is the optimal trajectory of M_k for the optimal switching strategy (s^{1*}, s^{2*}) . The black dots show the behavior of the optimal switching signal.

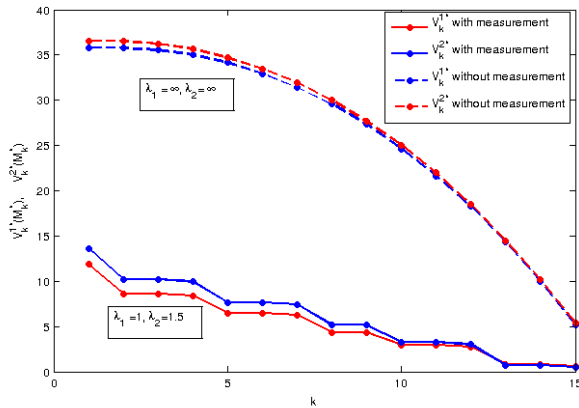


Fig. 3. A comparison among the costs for the cases when costly measurements are available ($\lambda_1 = 1, \lambda_2 = 1.5$) and no measurements are available ($\lambda_1 = \lambda_2 = \infty$)

We set the horizon of the game to be $T = 15$ and assume that X_0 is known to the players i.e. $M_0 = \mathbf{0}$.

In Figure 2, we show the optimal switching strategy Δ_k^* ($\equiv V_k^{1*} \cdot V_k^{2*}$) in black dots. In a game with horizon 15, the switch was closed for 5 times. In red line, we plot the value function $\mathcal{V}_k^{1*}(M_k^*)$ along the optimal trajectory of M_k determined by (19) and the optimal Δ_k^* . Similarly, in blue lines we plot $\mathcal{V}_k^{2*}(M_k^*)$.

In Figure 3, we illustrate a comparative result for the cases when observation costs are finite ($\lambda_1 = 1, \lambda_2 = 1.5$) and when observation costs are infinite (so that no observation is practically acquired). In this figure we see that even with 5 observations (out of 15 possible), there are more than 50% reductions in costs. The dotted curves in this figure also indicate the envelop of \mathcal{V}_k^{i*} . In other words, all the graphs of $\mathcal{V}_k^{i*}(M_k^*)$ obtained by varying the pair (λ_1, λ_2) will remain

below the dotted lines shown in Figure 3.

VI. CONCLUSION

In this work, we have considered a switched stochastic LQ game where the switching carries a finite cost. We have characterized the SPE control and switching strategies for both the players. The SPE control strategy turns out to be a linear strategy characterized by Riccati equations which do not depend on the switching strategy. The quality of state estimation depends on the switching strategy and hence the switching cost-to-go function depends on the estimation error variance. We have shown that no-switch is always a SPE. However, at certain time instances coordinated switching is also a SPE. Moreover when both no-switch and switch are SPE, then the cost-to-go with switching is lower than the same with no-switching for both players.

A valid question to ask is how a centralized agent would select its action strategies in order to optimize the welfare cost (i.e. the sum of two individual players' cost). The cost-to-go at stage k being the social welfare $\mathcal{V}_k = \mathcal{V}_k^1 + \mathcal{V}_k^2$. It is straightforward to notice $\mathcal{V}_k^* \leq \mathcal{V}_k^{1*} + \mathcal{V}_k^{2*}$.

The centralized switching strategy is determined by

$$\bar{C}_k(M, 0) + \mathcal{V}_{k+1}^*(AMA' + S) > \bar{C}_k(M, 1) + \mathcal{V}_{k+1}^*(\mathbf{0})$$

where $\bar{C}_k(\cdot, \cdot) = \bar{C}_k^1(\cdot, \cdot) + \bar{C}_k^2(\cdot, \cdot)$. An interesting study will be to characterize the social loss $l_k = \mathcal{V}_k^{1*} + \mathcal{V}_k^{2*} - \mathcal{V}_k^*$. This is known as price of anarchy, an interesting aspect to investigate for this problem.

We studied a two-player game, however similar analysis can be carried out for a general n -player game.

REFERENCES

- [1] T. Basar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1995, vol. 200.
- [2] J. Engwerda, *LQ dynamic optimization and differential games*. John Wiley & Sons, 2005.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] R. Isaacs, *Differential games: a mathematical theory with applications to warfare and pursuit, control and optimization*. Courier Corporation, 1999.
- [8] 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.
- [9] 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.
- [10] D. Maity and J. S. Baras, "Optimal strategies for stochastic linear quadratic differential games with costly information," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE, 2016, pp. 276–282.
- [11] D. Fudenberg and J. Tirole, *Game theory*. MIT Press, Cambridge, Massachusetts, 1991.
- [12] D. Maity, A. Anastasopoulos, and J. S. Baras, "Linear quadratic games with costly measurements," *ArXiv e-prints*, 2017.