Towards a Theory of Stochastic Adaptive Differential Games

Yan Li and Lei Guo

Academy of Mathematics and Systems Science,
Chinese Academy of Sciences

5th Swedish-Chinese Conference on Control, Lund, May 30-31, 2011

Outline

I. Introduction

II. Problem Formulation

III. Main Results

IV. Concluding Remarks

I. Introduction

Complex Systems and Game Theory

- **Complex systems with game-like relationships may be the most complicated ones to handle.**
  - **Politics, Economics, Business and Biology** *et al.*

    e.g., social choice theory, auctions, bargaining, evolutionary, some seemingly incongruous phenomena in nature such and cooperation and altruism ......
- **Game Theory appears to be a useful tool in modeling and analyzing conflicts in the context of dynamical systems.**

Differential Games

- **Motivated by combat problems and described by differential equations with payoff functions**
- **Combine game theory and control theory in some sense and related to optimal control closely**
  - **Two or more controls v.s. a single control**
  - **Each player has its own goal v.s. only one criterion to be optimized.**

An Example: Pursuer & Evader

- **The pursuer attempts to intercept the evader before some fixed time T while the latter attempts to do the opposite; both have limited energy sources.**
  - **e.g., a missile tracking down an airplane**
  - **The pursuer and the evader have opposite aims, one wants to minimize their distance while the other wants to maximize, just like the zero-sum game.**

An Example (mathematical description)

Determine a saddle point $(\boldsymbol{u(t; x_0, t_0), v(t; x_0, t_0)})$ for

$$J = \frac{a^2}{2}\|x_p(T) - x_e(T)\|^2_{A^T A} + \frac{1}{2}\int_{t_0}^{T}[\|u(t)\|^2_{R_p(t)} - \|v(t)\|^2_{R_e(t)}]\mathbf{d}t$$

subject to the constraints

$$\dot{x}_p = F_p(t)x_p + \bar{G}_p(t)u; \qquad x_p(t_0) = x_{p_0}$$

$$\dot{x}_e = F_p(t)x_e + \bar{G}_p(t)v; \qquad x_e(t_0) = x_{e_0}$$

and

$$u(t), v(t) \in R^m$$

An Example (cont'd)

- x_p **describes the state of the pursuer, while** x_e **describes the state of the evader.**
- a^2 **is introduced for weighting terminal miss against energy.**
- **A saddle point is defined as the pair** (u^0, v^0) **satisfying**

$$J(u^0, v) \leq J(u^0, v^0) \leq J(u, v^0)$$

for all $u, v \in R^m$**.**

Progress in Differential Games

- **Much progress has been made:**

  From zero-sum to nonzero-sum

  From deterministic to stochastic

  From perfect information to imperfect state information

- **Few adaptive results:**

  Few have considered adaptation issues in differential games where there are unknown parameters to the players. Partly because of the difficulty in the theoretical study of even the simpler LQG adaptive control problem.

II. Problem Formulation

Problem Formulation

- **The system is descried by**

$$\mathrm{d}X(t) = (AX(t) + B_1U_1(t) + B_2U_2(t))\mathrm{d}t + D\mathrm{d}W(t),$$

where $X(t) \in R^n$ denotes the ***state trajectory*** of the game.

$U_1(t) \in R^{m_1}$ is the ***strategy of Player 1.***

$U_2(t) \in R^{m_2}$ is the ***strategy of Player 2.***

$(W(t), \mathcal{F}_t; t \geq 0)$ is an R^p-valued standard Wiener process

B_1 and B_2 are unknown to both players.

Problem Formulation(cont'd)

- **The payoff function is**

$$J(U_1, U_2) = \overline{\lim_{T\to\infty}} \frac{1}{T} \int_0^T \big(X^T(t) Q X(t) + U_1^T(t) Q_1 U_1(t) - U_2^T(t) Q_2 U_2(t) \big) \mathrm{d}t,$$

where $Q = Q^T \geq 0$, $R_1 = R_1^T > 0$, $R_2 = R_2^T > 0$.

Player 1 aims to **minimize** the payoff function.

Player 2 aims to **maximize** the payoff function.

Some Definitions

- **Information pattern:**

  Let $\eta^i(t) = \{X(s), 0 \le s \le \epsilon_t^i\}$, $0 \le \epsilon_t^i \le t$, $i = 1, 2$,

  where $\eta^i(t)$ determines the state information gained by Player i at time t, and ϵ_t^i denotes the last time of Player i gaining his information, so Player i can only make strategy depending on $\eta^i(t)$.

  We say Player i's information pattern is

  ***open-loop* pattern:** if $\eta^i(t) = \{X(0)\}$

  ***closed-loop perfect state* pattern:**

  if $\eta^i(t) = \{X(s), 0 \le s \le t\}$

  ***feedback* pattern:** if $\eta^i(t) = \{X(t)\}$

Some Definitions(cont'd)

- **Feedback Nash equilibrium:**

  For the zero-sum linear-quadratic differential game with both players under the feedback pattern, a pair of strategies (U_1^0, U_2^0) constitutes a feedback Nash equilibrium if it satisfies

$$J(U_1^0, U_2) \leq J(U_1^0, U_2^0) \leq J(U_1, U_2^0).$$

  – Since the definition is defined under feedback pattern, U_i is a mapping:

$$\boldsymbol{U_i : \eta^i(t)(= X(t)) \rightarrow R^{m_i}}$$

The Standard Non-adaptive Case

- **The feedback Nash equilibrium for the above game is expressed as**

$$U_1(t) = -Q_1^{-1} B_1^T R X(t)$$
$$U_2(t) = Q_2^{-1} B_2^T R X(t),$$

where R is the symmetric solution of the following algebraic Riccati equation, which makes $A - (B_1 Q_1^{-1} B_1^T - B_2 Q_2^{-1} B_2^T) R$ stable

$$RA + A^T R + Q - R(B_1 Q_1^{-1} B_1^T - B_2 Q_2^{-1} B_2^T) R = 0$$

provided that some conditions are satisfied.

Assumptions

- **1) A is stable, and the pair $(A, [B_1, B_2])$ is controllable.**
- **2) The matrix function $G(s)$ is antianalytic perfactorizable,**

  where $G(s) = L + B^T(-sI - A^T)^{-1}Q(sI - A)^{-1}B$

  and $B = [B_1, B_2], L = \begin{bmatrix} Q_1 & \\ & -Q_2 \end{bmatrix}.$

Definitions

Assume $(A, [B_1, B_2])$ is stabilizable, and introduce a set

$$\mathcal{F}(A, B_1, B_2) \triangleq \Big\{ F \triangleq \begin{bmatrix} F_1, F_2 \end{bmatrix} \mid$$
$$A + B_1 F_1 + B_2 F_2 \text{ exponentially stable} \Big\}.$$

We say that $G(s)$ is *antianalytic perfactorizable* if there exists $F \in \mathcal{F}(A, B_1, B_2)$ such that $\widetilde{G}(s)$ defined below is antianalytic facotorizable:

$$\widetilde{G}(s) = L + B^T(-sI - \widetilde{A}^T)^{-1} F^T L + LF(sI - \widetilde{A})^{-1} B$$
$$+ B^T(-sI - \widetilde{A}^T)^{-1}(Q + L^T RL)(sI - \widetilde{A})^{-1},$$

which means that $\widetilde{G}(s)$ can be factorized into two proper rational matrix functions with their inverse also having the same property.

Remark

If Assumption 1) is relaxed to

1)′ The pair $(A, [B_1, B_2])$ is stabilizable,

then Assumptions 1)′ and 2) are **equivalent to** the property that the following algebraic Riccati equation

$$RA + A^T R + Q - R(B_1 Q_1^{-1} B_1^T - B_2 Q_2^{-1} B_2^T)R = 0$$

has a symmetric solution R, making

$$A - (B_1 Q_1^{-1} B_1^T - B_2 Q_2^{-1} B_2^T)R$$

stable (so the feedback Nash equilibrium exists).

When the Parameters are Unknown

- We use the certainty equivalence principle and estimate the players' unknown parameters B_1 and B_2 first.
- As a starting point, we assume that the two players use a common estimator, just like there is an independent agency providing parameter estimation or prediction for them.
- Because of the good convergence properties, we will use the weighted least squares(WLS) algorithms.

Linear Regression

- To put the system into a standard linear regression form, we introduce the following notations:

$$\theta^T = [B_1, B_2]$$

and

$$\varphi(t) = \begin{bmatrix} U_1(t) \\ U_2(t) \end{bmatrix}.$$

Then the system can be rewritten as

$$\mathrm{d}X(t) = \theta^T \varphi(t)\mathrm{d}t + D\mathrm{d}W(t).$$

WLS Algorithm

- The continuous-time WLS estimates, $\big(\theta(t), t \geq 0\big)$, can be defined by

$$\mathrm{d}\theta(t) = a(t)P(t)\varphi(t)[\mathrm{d}X^T(t) - X^T(t)A^T - \varphi^T(t)\theta(t)\mathrm{d}t],$$

$$\mathrm{d}P(t) = -a(t)P(t)\varphi(t)\varphi^T(t)P(t)\mathrm{d}t,$$

where $P(0) > 0$, $B_1(0)$ and $B_2(0)$ are arbitrary deterministic matrices such that the pair $\big(A, [B_1(0), B_2(0)]\big)$ is controllable.

The Choice of the Weights

- In order to guarantee the self-convergence property of WLS, the weights $a(t)$ is chosen like the following

$$a(t) = \frac{1}{f\big(r(t)\big)},$$

where $r(t) = \| P^{-1}(0) \| + \int_0^t U_1^T(s)U_1(s) + U_2^T(s)U_2(s)\mathrm{d}s$
and $f \in \mathbb{F}$ with

$$\mathbb{F} = \{f | f : \mathbb{R}_+ \to \mathbb{R}_+, f \text{ is slowly increasing}$$
$$\text{and} \int_c^\infty \frac{\mathrm{d}x}{xf(x)} < \infty \text{ for some } c \geq 0\},$$

where a function is called slowly increasing if it is increasing and satisfies $f \geq 1$.

Lemma 1:

The continuous-time WLS estimates $(\theta(t), t \geq 0)$ have the following properties:

$$1) \sup_{t \geq 0} |P^{-1}(t)\widetilde{\theta}(t)|^2 < \infty \quad \text{a.s. } ;$$

$$2) \int_0^{\infty} a(t)|\widetilde{\theta}^T(t)\varphi(t)|^2 \mathrm{d}t < \infty \quad \text{a.s.};$$

$$3) \lim_{t \to \infty} \theta(t) = \overline{\theta} \quad \text{a.s.};$$

for $i = 1, 2$, where $\widetilde{\theta}(t) = \theta(t) - \theta$, $\theta^T(t) = [B_1(t), B_2(t)]$ and $\overline{\theta}$ is a random matrix.

Remark

- **By Lemma 1, we know that the WLS algorithm is self-convergent, but the controllability of** $\Big(A, [B_1(t), B_2(t)]\Big)$ **is not guaranteed.**
- **This has also been the main difficulty encountered in the adaptive LQG control problem, which can been solved by using a random regularization method (see, Guo,IEEE-TAC, 1996; Duncan-Guo-Pasik Duncan,IEEE-TAC, 1999)**

Regularization

- By Lemma 1, we have

$$\| \theta - \theta(t) \| = O\Big(\| P(t) \| \Big).$$

- So we proceed to modify the estimates by the following way:

$$\theta(t, \beta) = \theta(t) - P^{1/2}(t)\beta,$$

where $\beta \in \mathcal{M}(m_1 + m_2, n)$, which denotes the family of $(m_1 + m_2) \times n$ real matrices, and we denote that

$$\theta^T(t, \beta) = [B_1(t, \beta), B_2(t, \beta)].$$

Uniform Controllability

- In order to show how to choose β, we will first state a definition for **uniformly controllable:**

  A family matrices $(A(t), B(t), A(t) \in R^{n\times n}, B(t) \in R^{n\times m}, t \geq 0)$ is said to be ***uniformly controllable*** if there is a constant $c > 0$ such that

$$\sum_{i=0}^{n-1} A^i(t)B(t)B^T(t)A^{iT}(t) \geq cI$$

  for all $t \in [0, \infty)$.

Choice of β

- The uniform controllability of $\big(A, [B_1(t,\beta), B_2(t,\beta)]\big)$, is equivalent to the uniform positivity of $F(t,\beta)$, where
$$F(t,\beta) = \det\Big(\sum_{k=0}^{n-1} A^k [B_1(t,\beta), B_2(t,\beta)] \begin{bmatrix} B_1^T(t,\beta) \\ B_2^T(t,\beta) \end{bmatrix} A^{kT}\Big).$$
- To ensure the uniform controllability of $\big(A, [B_1(t,\beta), B_2(t,\beta)]\big)$, β can be chosen like the following:

$$\beta_0 = 0$$
$$\beta_k = \begin{cases} \eta_k, & \text{if } F(k,\eta_k) \geq (1+\gamma)F(k,\beta_{k-1}) \\ \beta_{k-1}, & \text{otherwise} \end{cases}$$

where $(\eta_k, k \in \mathbb{N})$ are i.i.d. $\mathcal{M}(m1+m2;n)$-valued random variables that are independent of $(W(t);\ t \geq 0)$ and $\gamma \in (0, \sqrt{2}-1)$ is fixed.

Regularized Parameters

- The regularized parameters $[\bar{B}_1(k), \bar{B}_2(k)]$ are given by

$$\begin{bmatrix} \bar{B}_1^T(k) \\ \bar{B}_2^T(k) \end{bmatrix} = \begin{bmatrix} B_1^T(k) \\ B_2^T(k) \end{bmatrix} - P^{1/2}(k)\beta_k.$$

- The estimates are given by:

$$\hat{B}_1(t) = \bar{B}_1(k)$$

$$\hat{B}_2(t) = \bar{B}_2(k)$$

for $t \in (k, k+1]$, where $k \in \mathbb{N}$.

Lemma 2 (properties of the regularized estimates):

Let Assumptions 1) and 2) be satisfied for the game. Then for any admissible strategies $(U_1(t), U_2(t); t \geq 0)$, the family of regularized WLS estimates $(\hat{B}_i(t), t \geq 0, i = 1, 2)$ have the following properties:

1) **Self-convergence**, that is, $\hat{B}_i(t)$ converges a.s. to some finite random matrix as $t \to \infty$ for $i = 1, 2$.

2) The family $\left(A, [\hat{B}_1(t), \hat{B}_2(t)]\right)$ is **uniformly controllable**.

3) **Semiconsistency**, that is, as $t \to \infty$,

$$\int_0^t |(\hat{B}_i(s) - B_i)U_i(s)|^2 \mathrm{d}s = o(r(t)) + O(1) \text{ a.s.}$$

for $i = 1, 2$.

Remarks

- **By Lemma 2, we know that** $\left(A, [\hat{B}_1(t), \hat{B}_2(t)]\right)$ **is uniformly controllable with respect to** t.
- **We can also prove that,** $\hat{G}_t(s)$ **is antianalytic perfactorizable,**
  where $\hat{G}_t(s) = L + \hat{B}^T(-sI - A^T)^{-1}Q(sI - A)^{-1}\hat{B}$
  and $B = [\hat{B}_1(t), \hat{B}_2(t)], L = \begin{bmatrix} Q_1 & \\ & -Q_2 \end{bmatrix}$.

Adaptive Strategies

- Now, since $\left(A, [\hat{B}_1(t), \hat{B}_2(t)]\right)$ satisfies Assumptions 1) and 2), the following algebraic Riccati equation will have a real stable positive solution for each $t \in [0, \infty)$:

$$A^T R(t) + R(t)A + Q - R(t)\left(\hat{B}_1(t)Q^{-1}\hat{B}_1^T(t) - \hat{B}_2(t)Q^{-2}\hat{B}_2^T(t)\right)R(t) = 0.$$

- Then Player 1 can use the adaptive strategy given by

$$\boldsymbol{U_1(t) = -Q_1^{-1}\hat{B}_1^T(t)R(t)X(t)},$$

while the adaptive strategy for Player 2 is given by

$$\boldsymbol{U_2(t) = Q_2^{-1}\hat{B}_2^T(t)R(t)X(t)}.$$

III. Main Results

Theorem 1 (stability):

Let Assumptions 1) and 2) be satisfied and let the two players using the adaptive strategies as described above. Then the state trajectory $(X(t), t \geq 0)$ of the zero-sum linear-quadratic differential game is stable in the sense that

$$\limsup_{T \longrightarrow \infty} \frac{1}{T} \int_0^T |X(s)|^2 ds < \infty \quad \text{a.s.}$$

Strategies with Probing Signals

- To obtain the optimal strategy, diminishing probing signals are added to the adaptive strategies respectively, given by

$$U_1^*(t) = -Q_1^{-1}\hat{B}_1(t)R(k)X(t) + \gamma_k[V(t) - V(k)]$$

$$U_2^*(t) = Q_2^{-1}\hat{B}_2(t)R(k)X(t) + \gamma'_k[V'(t) - V'(k)]$$

for $t \in (k, k+1]$, $k \in \mathbb{N}$, and γ_k and γ'_k can be any sequences satisfying the following:

$$\frac{1}{k}\sum_{i=1}^{k}\gamma_i^2 = o(1), \ \ \log^l k = o(\sum_{i=1}^{k}\gamma_i^2) \text{ for any } l \geq 1$$

$$\frac{1}{k}\sum_{i=1}^{k}\gamma_i'^2 = o(1), \ \ \log^l k = o(\sum_{i=1}^{k}\gamma_i'^2) \text{ for any } l \geq 1$$

and where $V(t)$ and $V'(t)$ are sequences of independent standard Wiener Process that are independent of $(W(t); t \geq 0)$ and $(\eta_k; k \in \mathbb{N})$.

Theorem 2 (convergence):

Let Assumptions 1) and 2) be satisfied and let the players use the adaptive strategies with probing signals. Then estimates are consistent:

$$\lim_{t\to\infty} \hat{B}_1(t) = B_1 \qquad \text{a.s.}$$

$$\lim_{t\to\infty} \hat{B}_2(t) = B_2 \qquad \text{a.s.}$$

Theorem 3 (optimality):

The above defined pair of adaptive strategies (U_1^*, U_2^*) is a feedback Nash equilibrium for the following payoff function:

$$J(U_1, U_2)$$

$$= \limsup_{T\to\infty} \frac{1}{T}\int_0^T [X^T(t)QX(t) + U_1^\tau(t)R_1U_1(t) + U_2^\tau(t)R_2U_2(t)]\mathrm{d}t$$

i.e., for any pair of strategies (U_1, U_2) , it holds that

$$J(U_1^*, U_2) \le J(U_1^*, U_2^*) \le J(U_1, U_2^*).$$

IV. Concluding Remarks

Concluding Remarks

- This talk has discussed a class of linear quadratic two-player zero-sum stochastic differential games with unknown parameters, and has demonstrated that the optimality of the payoff function can be achieved by adaptive strategies.
- Many problems remain open, which includes the relaxation of the related conditions, the use of different estimators by different players, and the problems of many players in non-zero-sum differential games, and so on.

THANK YOU!