Two-player Ad hoc Output-feedback Cumulant Game Control

Chukwuemeka Aduba[1] and Chang-Hee Won[2]
[1]*Arris Group Inc., Horsham, PA 19044, U.S.A.*
[2]*Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, U.S.A.*

Keywords: Cost Cumulant Game Control, Nash Game, Neural Networks, Output-feedback, Statistical Control.

Abstract: This paper studies a finite horizon output-feedback game control problem where two players seek to optimize their system performance by shaping the distribution of their cost function through cost cumulants. We consider a two-player second cumulant nonzero-sum Nash game for a partially-observed linear system with quadratic cost function. We derive the near-optimal players strategy for the second cost cumulant function by solving the Hamilton-Jacobi-Bellman (HJB) equation. The results of the proposed approach are demonstrated by solving a numerical example.

1 INTRODUCTION

Game theory is the study of tactical interactions involving conflicts and cooperations among multiple decision makers called players with applications in diverse disciplines such as management, communication networks, electric power systems and control (Zhu et al., 2012), (Charilas and Panagopoulos, 2010), (Cruz et al., 2002). Stochastic differential game results from strategic interactions among players in a random dynamic system (Basar, 1999). In stochastic optimal control, there is a player and cost function to be optimized while in stochastic differential games, there are multiple players and separate cost function to be optimized by each player.

In most practical control engineering applications, not all the states are measurable. The system model may consists of unknown disturbances usually expressed as process noise while the inaccuracies in measurement are usually expressed as measurement noise. An approach to account for the unmeasurable states is to estimate those states using an estimator before utilizing the states in a controller in a feedback control system. This approach is part of a generalized method to analyzing linear stochastic systems by applying the concept of certainty equivalence principle (Van De Water and Willems, 1981) or related separation principle (Wonham, 1968). Bensoussan *et al.* (Bensoussan and Schuppen, 1985) investigated the stochastic optimal control problem for partially-observed system with exponential cost criterion and proved that separation theorem does not hold for such scenario. (Zheng, 1989) investigated both optimal and suboptimal approach to output feedback control for a linear system with quadratic cost function while the solvability of the necessary and sufficient conditions for the existence of a stabilizing output feedback solution for a continuous-time linear systems was studied in (Geromel et al., 1998). Aberkane *et al.* (Aberkane et al., 2008) investigated the output feedback solution for generalized stochastic hybrid linear systems and provided a dynamic system practical example. The infinite-horizon output feedback Nash game for a stochastic weakly-coupled system with state-dependent noise was studied in (Mukaidani et al., 2010). In addition, the necessary conditions for the existence of Nash equilibrium were given in (Mukaidani et al., 2010). Klompstra (Klompstra, 2000), extended risk-sensitive control to discrete time game theory and solved the Nash equilibrium for the partially observed state of a 2-player game.

In this paper, we are motivated to extend the above-referenced studies by considering higher-order statistics of cost function. In particular, we consider a second cumulant nonzero-sum Nash game for a partially-observed system of two players on a fixed time interval where the players shape the distribution of their cost cumulant function to improve system performance. This form of dynamic game finds application in satellite and mobile robot systems. The second cumulant of cost function is equivalent to the variance of the cost function. However, the optimization of cost function distribution through cost cumulant was initiated by Sain (Sain, 1966), (Sain and Liberty,

Aduba C. and Won C..
Two-player Ad hoc Output-feedback Cumulant Game Control.
DOI: 10.5220/0005503300530059
In *Proceedings of the 12th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2015), pages 53-59
ISBN: 978-989-758-122-9

1971) while Won *et al.* (Won et al., 2010), extended the theory of cost cumulant to second, third and fourth cumulants for a nonlinear system with non-quadratic cost and derived the corresponding HJB equations.

The reminder of this paper is organized as follows. In Section 2, we state the mathematical preliminaries and formulate the second cumulant game problem. Section 3 states the necessary condition for the existence of Nash equilibrium solution while Section 4 derives the players strategy based on solving the coupled Hamilton-Jacobi-Bellman (HJB) equations which is the main result of this paper. Section 5 describes the numerical approximate method for solving the coupled HJB equations while a numerical example is solved in Section 6. Finally, the conclusions are given in Section 7.

2 PROBLEM FORMULATION

Consider a 2-player linear state dynamics and measured output described by the linear Itô-sense stochastic differential equation.

$$\begin{aligned} dx(t) &= A(t)x(t)dt + \sum_{k=1}^{2} B_k(t)u_k(t)dt + G(t)dw_1(t), \\ dy(t) &= C(t)x(t)dt + D(t)dw_2(t), \end{aligned} \tag{1}$$

where $t \in [t_0, t_F] = T$, $x(t) \in \mathbb{R}^n$ is the state, $u_k(t) \in U_k \subset \mathbb{R}^m$ is the k-th player strategy, $k = 1,2$ and $w_1(t)$, $w_2(t)$ are Gaussian random process defined on a probability space (Ω_0, F, P) where Ω_0 is a nonempty set, F is a σ-algebra of Ω_0 and P is a probability measure on $(\Omega_0, F).x(t_0) = x_0$ is the initial state vector with covariance matrix P_0. The Gaussian random process $w_1(t)$ has zero mean and covariance of $E(dw_1(t)dw_1'(t)) = W_1(t)dt$ and similarly the Gaussian random process $w_2(t)$ has zero mean and covariance of $E(dw_2(t)dw_2'(t)) = W_2(t)dt$. The noise processes $w_1(t)$ and $w_2(t)$ are assumed independent with $E(dw_1(t)dw_2'(t)) = E(dw_2(t)dw_1'(t)) = 0$ assuming $dw_1, \underline{d}w_2$ have same dimension. Let $\underline{Q_0} = [t_0, t_F) \times \mathbb{R}^n$, $\overline{Q}_0$ denote the closure of Q_0, i.e $\overline{Q}_0 = T \times \mathbb{R}^n$. Assume there exist constants $c_1, c_2 > 0 \in \mathbb{R}$ such that

$$\|A(t)\| + \sum_{k=1}^{2} \|B_k\| \le c_1, \|G(t)\| \le c_2, \tag{2}$$

where $A(.), B_k(.), C(.), D(.), G(.)$ are elements of $C^1([t_0, t_F])$ with appropriate dimensions. Let a feedback strategy law be defined as $u_k(t) = \mu_k(t, x(t)), t \in T$. Then, (1) can be written as

$$dx(t) = f^{\mu_k}(x(t))dt + G(t)dw_1(t), x(t_0) = x_0, \tag{3}$$

where $f^{\mu_k}(x)$ denotes $A(t)x(t) + \sum_{k=1}^{2} B_k(t)u_k(t)$. There exist a bounded, borel measurable feedback strategy $\mu_k(x) : \mathbb{R}^m \to U_k$ such that $\mu_k(x)$ satisfies a global Lipschitz condition: i.e there exists a constant c_1 such that

$$\|\mu_k(x_1) - \mu_k(x_2)\| \le c_1 \|x_1 - x_2\|, \tag{4}$$

$\|.\|$ is the Euclidean norm and $x_1, x_2 \in \mathbb{R}^n$. Also, $\mu_k(x)$ satisfies linear growth condition

$$\|\mu_k(x)\| \le c_2(1 + \|x\|). \tag{5}$$

Then, if $E\|x(t)\|^2$ is finite, there is a unique solution to (1) which is a Markov diffusion process on $\mathbb{R}^n$ (Fleming and Rishel, 1975). In order to assess performance of (1), consider the cost function (J^k) for the k-th player given as:

$$\begin{aligned} &J^k(t_0, x(t_0), \mu_1, \mu_2) = x'(t_F)Q_f x(t_F) + \\ &\int_{t_0}^{t_F} \Big[x'(s)Q(s)x(s) + \sum_{i=1}^{2} \mu_i'(s)R_{ki}(s)\mu_i(s) \Big] ds, \end{aligned} \tag{6}$$

where $k = 1,2$, $x(t_F) = x_f$, $Q(s), Q_f$ are symmetric positive semi-definite and $R_{ki}(.)$ is symmetric positive definite, which can also be represented as

$$J^k(t_0, x(t_0), \mu_1, \mu_2) = \int_{t_0}^{t_F} L^k(s, x, \mu_1, \mu_2) ds + \psi^k(x(t_F)), \tag{7}$$

where $k = 1,2$, L^k is the running cost, ψ^k is the terminal cost and L^k, ψ^k both satisfy polynomial growth condition. Let the state estimate be $\hat{x}(t)$ and the state estimate error be $\overline{x}(t)$ where $x(t)$ is the state true value. Then, the state estimation error $\overline{x}(t)$, is given as

$$\overline{x}(t) = x(t) - \hat{x}(t). \tag{8}$$

The filtered state estimate $\hat{x}(t)$ is given as

$$\begin{aligned} d\hat{x}(t) =& A(t)\hat{x}(t)dt + \sum_{k=1}^{2} \Big(B_k(t)u_k(t) \Big) dt \\ &+ K(t) \Big(dy(t) - C(t)\hat{x}(t)dt \Big). \end{aligned} \tag{9}$$

where $K(t)$ is the Kalman Filter gain (Davis, 1977).

Lemma 2.1. The expected value of the cost function (6) conditioned on the σ-algebra generated by the measured output (1) can be rewritten as

$$\begin{aligned} E\Big\{ J^k(t_0, \hat{x}(t_0), \mu_1, \mu_2) \Big\} &= \int_{t_0}^{t_F} \Big[E\Big(\hat{x}'(s)Q(s)\hat{x}(s) \Big) \\ &+ \mathrm{tr}\Big(Q(s)P(s) \Big) \Big] ds + \int_{t_0}^{t_F} \Big[\sum_{i=1}^{2} \mu_i'(s)R_{ki}(s)\mu_i(s) \Big] ds \\ &+ E\left\{ \hat{x}'(t_F)Q_f\hat{x}(t_F) \right\} + \mathrm{tr}\Big(Q_f P_f \Big), \end{aligned} \tag{10}$$

where $k=1,2$, $\hat{x}(t_F)=\hat{x}_f$, $Q(.),Q_f,P(.),P_f$ are positive semi-definite, $R_{ki}(s)$ is positive definite for $k=i$ and positive semi-definite for $k\neq i$, $P(.),P_f$ are the state error estimate covariances.

Proof. See (Davis, 1977) for single player case, a two-player case follows similar derivation.

Furthermore, we utilize the backward evolution operator, $O^k(\mu_1,\mu_2)$, as defined in (Sain et al., 2000): $O^k(\mu_1,\mu_2)=O_1^k(\mu_1,\mu_2)+O_2^k(\mu_1,\mu_2)$,

$$\begin{aligned} O_1^k(\mu_1,\mu_2) &= \frac{\partial}{\partial t}+f'(t,x,\mu_1,\mu_2)\frac{\partial}{\partial x}, \\ O_2^k(\mu_1,\mu_2) &= \frac{1}{2}\mathrm{tr}\left(G(t)W_1(t)G(t)'\frac{\partial^2}{\partial x^2}\right), \end{aligned} \tag{11}$$

with tr = trace in (11). To study the cumulant game of cost function, the m-th moments of cost functions M_m^k of the k-th player is defined as:

$$M_m^k(t,\hat{x},\mu_1,\mu_2)=E\left\{(J^k)^m(t,x,\mu_1,\mu_2)|x(t)=x\right\}, \tag{12}$$

where $m=1,2$. The m-th cost cumulant function $V_m^k(t,\hat{x})$ of the k-th player is defined by (Smith, 1995),

$$V_m^k(t,\hat{x})=M_m^k-\sum_{i=0}^{m-2}\frac{(m-1)!}{i!(m-1-i)!}M_{m-1-i}^kV_{i+1}^k, \tag{13}$$

where $t\in T=[t_0,t_F]$, $x(t_0)=x_0$, $\hat{x}(t)\in\mathbb{R}^n$. Next, we introduce some definitions.

Definition 2.1. A function $M_1^k:\bar{Q}_0\rightarrow\mathbb{R}^+$ is an admissible first moment cost function if there exists a strategy μ_k such that

$$M_1^k(t,\hat{x})=M_1^k(t,\hat{x};\mu_1,\mu_2), \tag{14}$$

for $t\in T,\hat{x}\in\mathbb{R}^n$, $M_1^k\in C^{1,2}(\bar{Q}_0)$. Also, V_1^k is the admissible first cumulant cost function for the k-th player related to the moment function through the moment-cumulant relation (13). In addition, $\mu_k\in U_{M^k}, V_1^k(t,\hat{x})=V_1^k(t,\hat{x};\mu_1,\mu_2)$.

Definition 2.2. A class of admissible strategy U_{M^k} is defined such that if $\mu_k\in U_{M^k}\subset\mathbb{R}^m$ then μ_k satisfies the equality of Definition 2.1 for M_0^k,M_1^k. It should be noted that first moment M_1^k is the same as first cumulant V_1^k, $M_0^k=1$ and $V_0^k=0$.

Definition 2.3. Let V_1^k be the k-th player admissible cumulant cost functions. The player strategy μ_k^* is the k-th player equilibrium solution if it is such that

$$\begin{aligned} V_2^{1*}(t,\hat{x}) &= V_2^1(t,\hat{x};\mu_1^*,\mu_2^*)\leq V_2^1(t,\hat{x};\mu_1^*,\mu_2), \\ V_2^{2*}(t,\hat{x}) &= V_2^2(t,\hat{x};\mu_1^*,\mu_2^*)\leq V_2^2(t,\hat{x};\mu_1,\mu_2^*). \end{aligned} \tag{15}$$

for all $\mu_k\in U_{M^k}$ where the set $\{\mu_1^*,\mu_2^*\}$ is a Nash equilibrium solution and the set $\{V_2^{1*},V_2^{2*}\}$ is the Nash equilibrium value set.

Problem Definition. Consider an open set $Q\subset Q_0$ and let the k-th player cost cumulant functions $V_1^k(t,\hat{x})\in C_p^{1,2}(Q)\cap C(\bar{Q})$ be an admissible cumulant function where the set $C_p^{1,2}(Q)\cap C(\bar{Q})$ means that the function V_1^k satisfy polynomial growth condition and is continuous in the first and second derivatives of Q, and continuous on the closure of Q. Assume the existence of a near-optimal strategy $\mu_k^*\in U_{M^k}$ and near-optimal value function $V_2^{k*}(t,\hat{x})\in C_p^{1,2}(Q)\cap C(\bar{Q})$ for the k-th player. Thus, a 2-player second cumulant output feedback game problem is to find the Nash strategy $\mu_k^*(t,\hat{x})$ for the partially-observed linear state system with $k=1,2$ which results in the near-optimal 2^{nd} value function $V_2^{k*}(t,\hat{x})$ given as

$$\begin{aligned} V_2^{1*}(t,\hat{x}) &= \min_{\mu_1\in U_{M^1}}\left\{V_2^1(t,\hat{x};\mu_1,\mu_2)\right\}, \\ V_2^{2*}(t,\hat{x}) &= \min_{\mu_2\in U_{M^2}}\left\{V_2^2(t,\hat{x};\mu_1,\mu_2)\right\}. \end{aligned} \tag{16}$$

Remarks. To find the Nash equilibrium strategies $\mu_1^*(t,\hat{x}),\mu_2^*(t,\hat{x})$, we constrain the candidates of the near-optimal players strategy to U_{M^1},U_{M^2}, and the near-optimal value functions $V_2^{1*}(t,\hat{x}),V_2^{2*}(t,\hat{x})$ are found with the assumption that $V_1^1(t,\hat{x}),V_1^2(t,\hat{x})$, are admissible.

3 AD HOC OUTPUT FEEDBACK CUMULANT GAME

Theorem 3: From the full-state feedback statistical control in (Won et al., 2010), the minimal 2^{nd} value function $V_2^{k*}(t,x)$ for (1) with zero measurement noise satisfies the following HJB equation for the k-th player:

$$\begin{aligned} 0=\min_{\mu_k\in U_{M^k}}&\left\{O^k(\mu_1^*,\mu_2^*)\left[V_2^{k*}(t,x)\right]\right. \\ &\left.+\left(\frac{\partial V_1^k(t,x)}{\partial x}\right)'G(t)W_1G(t)'\left(\frac{\partial V_1^k(t,x)}{\partial x}\right)\right\}, \end{aligned} \tag{17}$$

with the terminal condition $V_j^k(t_F,x_F)=0$, $k=1,2$, $j=1,2$. Assuming separation principle (Wonham, 1968), the minimal 2^{nd} value function $V_2^{k*}(t,\hat{x})$ for (9) satisfies the following HJB equation for the k-th player:

$$0 = \min_{\mu_k \in U_{M^k}} \left\{ O^k(\mu_1^*, \mu_2^*)\left[V_2^{k*}(t,\hat{x})\right] + \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right)' K(t)W_2K(t)' \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right) \right\}, \tag{18}$$

with terminal condition $V_j^k(t_F, \hat{x}_F) = 0$, $k = 1,2$, $j = 1,2$, $K(t)$ is the Kalman filter gain associated with (1) after transformation through innovative process (Kailath, 1968), (Davis, 1977).

Remark. The HJB equation (17) provides a necessary condition for the existence of equilibrium solution of a 2-player, 2^{nd} cost cumulant game. A similar condition with proof is given for statistical control in (Won et al., 2010). Our approach in (18) is termed ad hoc, since we assume that separation principle holds for the stochastic linear system with 2^{nd} cumulant function $V_2^k(t,\hat{x})$.

4 TWO-PLAYER CUMULANT GAME NASH STRATEGY

Theorem 4. Let the solution to the k-th player second cumulant output feedback game be given by

$$\mu_k^*(t,\hat{x}) = -\frac{1}{2}R_{kk}^{-1}B_k'\left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}} + \gamma_{2k}\frac{\partial V_2^{k*}(t,\hat{x})}{\partial \hat{x}}\right), \tag{19}$$

where γ_{2k} is the Lagrange multiplier and V_1^k, V_2^k are the first, second cumulant cost functions and solutions of the following coupled HJB equations:

$$\begin{aligned} 0 =& O^k(\mu_{-k},\mu_k)\left[V_1^k(t,\hat{x})\right] + M_0^k(t,\hat{x})L^k(t,\hat{x},\mu_{-k},\mu_k), \\ 0 =& O^k(\mu_{-k},\mu_k)\left[V_2^k(t,\hat{x})\right] \\ &+ \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right)' K(t)W_2K(t)' \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right), \end{aligned} \tag{20}$$

where $M_0^k = 1$, $O^k(.)$ is the backward operator, $-k$ represents not k; if k is 1 then $-k$ is 2 and vice-versa.

Proof. From the system equation (1), (18) and assuming that separation principle holds, the minimal 2^{nd} value function $V_2^{k*}(t,\hat{x})$ satisfies the following HJB equation for the k-th player.

$$0 = \min_{\mu_k \in U_{M^k}} \left\{ O^k(\mu_1^*, \mu_2^*)\left[V_2^{k*}(t,\hat{x})\right] + \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right)' K(t)W_2K(t)' \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right) \right\}, \tag{21}$$

with terminal condition $V_j^k(t_F, \hat{x}_F) = 0$, $k = 1,2$, $j = 1,2$, $K(t)$ is the Kalman filter gain. Since the first cost cumulant function V_1^k is admissible (def. 2.1), the following coupled equations are satisfied

$$\begin{aligned} 0 =& O^k(\mu_{-k},\mu_k)\left[V_1^k(t,\hat{x})\right] + M_0^k(t,\hat{x})L^k(t,\hat{x},\mu_{-k},\mu_k), \\ 0 =& O^k(\mu_{-k},\mu_k)\left[V_2^k(t,\hat{x})\right] \\ &+ \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right)' K(t)W_2K(t)' \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right), \end{aligned} \tag{22}$$

where $M_0^k = 1$, $O^k(.)$ is the backward operator and the first line of (22) follows from the classical HJB equation while the second line relates the second cumulant function with the first cumulant function in the HJB equation. Thus, converting (22) to unconstrained optimization problem gives

$$\begin{aligned} 0 = \min_{\mu_k \in U_{M^k}} \Bigg\{ & O^k(\mu_{-k},\mu_k)\left[V_1^k(t,\hat{x})\right] \\ &+ M_0^k(t,\hat{x})L^k(t,\hat{x},\mu_{-k},\mu_k) + \gamma_{2k}(t)O^k(\mu_{-k},\mu_k)\left[V_2^{k*}(t,\hat{x})\right] \\ &+ \gamma_{2k}(t)\left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right)' K(t)W_2K(t)' \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right) \Bigg\}, \end{aligned} \tag{23}$$

where γ_{2k} is the Lagrange multiplier. From backward operator (11) using (9), (10) and expanding (23) gives

$$\begin{aligned} \min_{\mu_k \in U_{M^k}} \Bigg\{ & \left(\frac{\partial V_1^k(t,\hat{x})}{\partial t}\right) + \hat{x}'(t)Q(t)\hat{x}(t) \\ &+ \mathrm{tr}\Big(Q(t)P(t)\Big) + \sum_{i=1}^{2}\mu_i(t)R_{ki}(t)\mu_i'(t) \\ &+ \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right)\left(\hat{x}(t)'A(t)' + \sum_{i=1}^{2}\mu_i(t)'B_i(t)'\right) \\ &+ \frac{1}{2}\mathrm{tr}\left(K(t)W_2K(t)'\left(\frac{\partial^2 V_1^k(t,\hat{x})}{\partial \hat{x}^2}\right)\right) + \gamma_{2k}\left(\frac{\partial V_2^{k*}(t,\hat{x})}{\partial t}\right) \\ &+ \gamma_{2k}\left(\frac{\partial V_2^{k*}(t,\hat{x})}{\partial \hat{x}}\right)\left(\hat{x}(t)'A(t)' + \sum_{i=1}^{2}\mu_i(t)'B_i(t)'\right) \\ &+ \frac{\gamma_{2k}}{2}\mathrm{tr}\left(K(t)W_2K(t)'\left(\frac{\partial^2 V_2^{k*}(t,\hat{x})}{\partial \hat{x}^2}\right)\right) \\ &+ \gamma_{2k}\left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right)' K(t)W_2K(t)' \left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}}\right) \Bigg\} = 0. \end{aligned} \tag{24}$$

Minimizing (24) with respect to $\mu_k(t,\hat{x})$ gives

$$\mu_k^*(t,\hat{x}) = -\frac{1}{2}R_{kk}^{-1}B_k'\left(\frac{\partial V_1^k(t,\hat{x})}{\partial \hat{x}} + \gamma_{2k}\frac{\partial V_2^{k*}(t,\hat{x})}{\partial \hat{x}}\right). \tag{25}$$

Remark. The strategy for the k-th player $\mu_k^*(t,\hat{x})$ derived from the coupled HJB equation (22) is suboptimal. In (22), the certainty equivalent principle has been extended to the second cumulant output feedback game where a Kalman filter is used for state estimation.

5 NUMERICAL APPROXIMATION METHOD

The analytical solutions of HJB equation (18) is difficult to find except for simple linear systems. Sanberg (Sandberg, 1998) showed that neural networks (NN) with time-varying weights can be utilized to approximate uniformly continuous time-varying functions. We are motivated by the work in (Chen et al., 2007), to extend NN approach to cost cumulant game. In this approach, NN is utilized to approximate the value function based on method of least squares on a pre-defined region. The value functions V_i^k can be approximated as $V_i^k(t,\hat{x}) = \mathbf{w}_L'(t)\Lambda_L(\hat{x}) = \sum_{i=1}^{L} w_i(t)\gamma_i(\hat{x})$ on t on a compact set $\Omega \to \mathbb{R}^n$. Thus, we approximate the players value functions V_m^k as $V_{mL}^k(t,\hat{x}) = \mathbf{w}_{mkL}'(t)\Lambda_{mkL}(\hat{x}) = \sum_{i=1}^{L} w_{mki}(t)\gamma_{mki}(\hat{x})$, where $\mathbf{w}_{mkL}(t)$ and $\Lambda_{mkL}(\hat{x})$ are vectors, and $\mathbf{w}_{mkL}(t) = \{w_{mk1}(t),\ldots,w_{mkL}(t)\}'$ and $\Lambda_{mkL}(\hat{x}) = \{\gamma_{mk1}(\hat{x}),\ldots,\gamma_{mkL}(\hat{x})\}'$ are the vector neural network weights and vector of activation functions and L is the number of the hidden-layer neurons. Using the approximated value functions $V_{mL}^k(t,\hat{x})$ in the HJB equations result in residual error equations. We apply weighted residual method (Finlayson, 1972) to minimize the residual error equations and then numerically solve for the least square NN weights (Chen et al., 2007).

6 SIMULATION

Consider a linear deterministic dynamic system in (Zheng, 1989), where we introduce gaussian noise as process and measurement noise. The stochastic system is represented as

$$\begin{aligned} dx(t) &= Ax(t)dt + B_1u_1(t)dt + B_2u_2(t)dt + Gdw_1(t),\\ dy(t) &= Cx(t)dt + Ddw_2(t), \end{aligned} \tag{26}$$

$$A = \begin{bmatrix} -2 & 0 & 3 & 2 \\ 4 & -2 & 1 & 3 \\ 2 & 3 & -3 & 4 \\ 0 & 0 & 0 & -2 \end{bmatrix}, B_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

$$B_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 & 0.1 \\ 0 & 1 & 0 & 0.1 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

and the state variable $x(t)$ is defined as: $x(t) = [x_1(t)\ x_2(t)\ x_3(t)\ x_4(t)]'$. We assume that G and D in (26) are 4×1 and 3×1 constant vectors given as $G = [1\ 1\ 1\ 1]'$, $D = [1\ 1\ 1]'$ and $dw_1(t), dw_2(t)$ in (26) as a Gaussian process with mean $E\{dw_1(t)\} = E\{dw_2(t)\} = 0$ and covariance $E\{dw_1(t)dw_1(t)'\} = 0.1$ and $E\{dw_2(t)dw_2(t)'\} = 0.1$. In this example, we study a 2-player 2^{nd} cumulant ad hoc output feedback Nash game. Here, we compute the suboptimal solution for the player strategy through solving the output feedback 2^{nd} cumulant game problem constraint on the 1^{st} cumulant cost function.

The first player cost function J^1 is

$$\begin{aligned} J^1(t_0,x(t_0),u_1(t_0),u_2(t_0)) &= \int_{t_0}^{t_F} \{x_1^2(t) + x_2^2(t)\\ &+ x_3^2(t) + x_4^2(t) + u_1^2(t)\}dt + \psi^1(x(t_F),t_F), \end{aligned} \tag{27}$$

where $\psi^1(x(t_F),t_F) = 0$ is the terminal cost and the second player cost function J^2 is

$$\begin{aligned} J^2(t_0,x(t_0),u_1(t_0),u_2(t_0)) &= \int_{t_0}^{t_F} \{x_1^2(t) + x_2^2(t)\\ &+ x_3^2(t) + x_4^2(t) + u_2^2(t)\}dt + \psi^2(x(t_F),t_F), \end{aligned} \tag{28}$$

where $\psi^2(x(t_F),t_F) = 0$ is the terminal cost. The activation function $\Lambda_L(x)$ for the value functions of the players are the same and based on (Chen and Jagannathan, 2008) which are formulated as

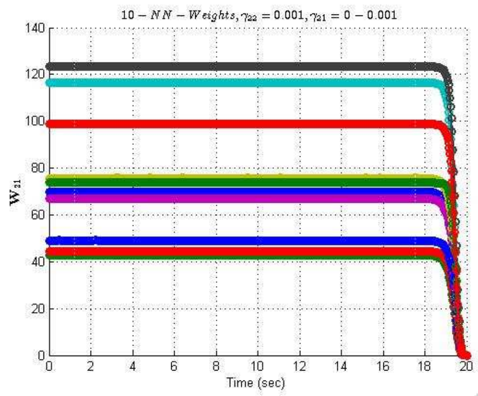
$$\Lambda_L(x) = \sum_{i=1}^{\frac{M}{2}} \left(\sum_{j=1}^{n} x_j\right)^{2i}, \tag{29}$$

where in (29), M is an even-order of the approximation, L is the number of hidden-layer neurons, n is the system dimension.

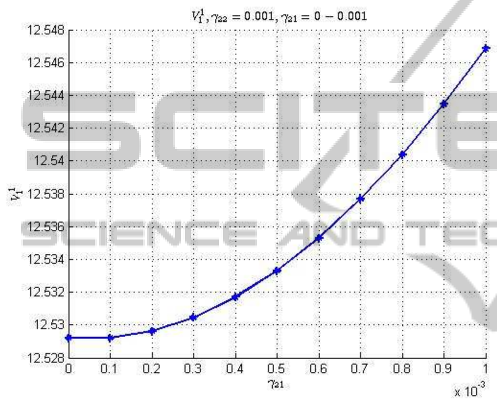
The input function $\Lambda_L(x)$ (29) is

$$\Lambda_L = \{x_1^2, x_2^2, x_3^2, x_4^2, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4\}'. \tag{30}$$

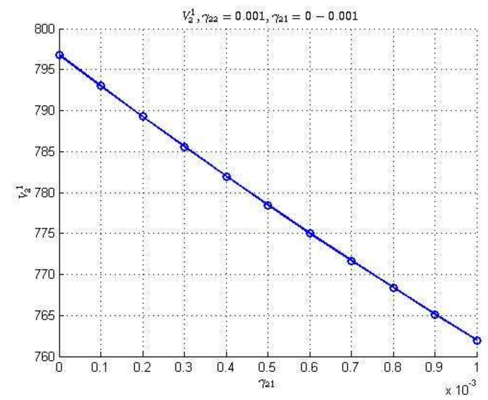
We transform this problem as an innovative process (Kailath, 1968) in terms of state estimate using (8), (9), (10) and solve for the Kalman filter gain. For the NN series approximation, we choose a polynomial function (30) of up to second-order ($M = 2$) in state variable (i.e x is 2^{nd} order) with length $L = 10$. Higher order polynomial did not provide significant improvement in the approximation accuracy. In the simulation, the asymptotic stability region for states was arbitrarily chosen as $-5 \le x_1 \le 5, -5 \le x_2 \le 5, -5 \le$



(a) NN Weights versus time

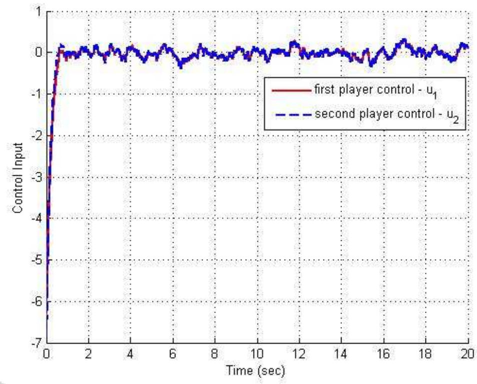


(b) V_1^1 versus γ_{21}

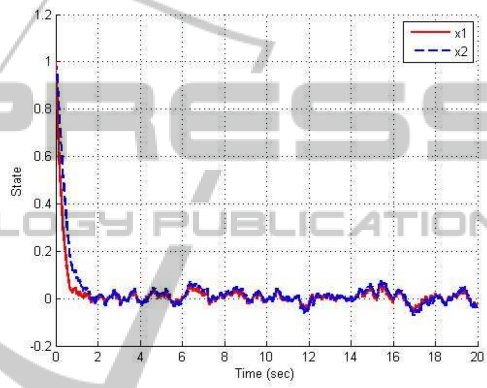


(c) V_2^1 versus γ_{21}

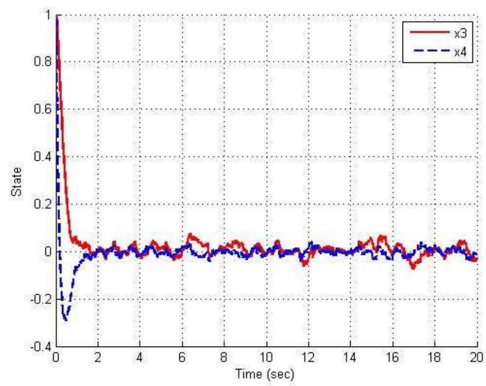
Figure 1: Neural Network Weights and Value Function.



(a) Control trajectory



(b) State trajectory



(c) State trajectory

Figure 2: Control Strategy and State Trajectory.

$x_3 \leq 5$ and $-5 \leq x_4 \leq 5$. The final time t_F was 20 s and $\mathbf{w}'_{11L}(t_F) - \mathbf{w}'_{21L}(t_F) = \{0\}$ and $\mathbf{w}'_{12L}(t_F) - \mathbf{w}'_{22L}(t_F) = \{0\}$. The initial condition was $x(t_0) = x_0 = [1\ 1\ 1\ 1]'$.

Figs. 1(a) to 1(c) show the first player neural network weights and value functions which are similar to the second player, hence only the first player plots are shown. Fig. 1(a), the neural network weights converge to constants. Plots 1(b) to 1(c) show the first and second value cumulant functions. From Fig. 1(b), it was observed that the value function V_1^1 increases with increase in γ_{21} while from Fig. 1(c), it was observed that the value functions V_2^1 decreases as γ_{21} increase. The Lagrange multipliers γ_{2k} were selected as constants. The Nash suboptimal controls, u_1 and u_2 are shown in Fig. 2(a). It should be noted from Fig. 2(a), that the Nash suboptimal controls for the two players were solved for the 2^{nd} cumulant game by selecting γ_{21}, γ_{22} where the value functions are minimum which in our case were $\gamma_{21} = 0.001$, $\gamma_{22} = 0.001$. In addition, we have the design freedom in γ_{2k} values selection to enhance system performance. From Figs. 2(b) to 2(c), the states converge to values close to zero.

7 CONCLUSIONS

In this paper, we analyzed an output feedback cumulant differential game control problem using cost cumulant optimization approach. We investigated a linear stochastic system with two players and derived a 2-player near-optimal strategies for the tractable auxiliary problem. The efficiency of our proposed method has been demonstrated using a numerical example where a neural network series method was applied to solve the HJB equations.

REFERENCES

Aberkane, S., Ponsart, J. C., Rodrigues, M., and Sauter, D. (2008). Output feedback control of a class of stochastic hybrid systems. *Automatica*, 44:1325–1332.

Basar, T. (1999). Nash Equilibria of Risk-Sensitive Nonlinear Stochastic Differential Games. *Journal of Optimization Theory and Applications*, 100(3):479–498.

Bensoussan, A. and Schuppen, J. H. V. (1985). Optimal Control of Partially Observable Stochastic Systems with an Exponential-of-Integral Performance Index. *SIAM Journal of Control and Optimization*, 23(4):599–613.

Charilas, D. E. and Panagopoulos, A. D. (2010). A survey on game theory applications in wireless networks. *Computer Networks*, 54(18):3421–3430.

Chen, T., Lewis, F. L., and Abu-Khalaf, M. (2007). A Neural Network Solution for Fixed-Final Time Optimal Control of Nonlinear Systems. *Automatica*, 43(3):482–490.

Chen, Z. and Jagannathan, S. (2008). Generalized Hamilton-Jacobi-Bellman Formulation-Based Neural Network Control of Affine Nonlinear Discrete-Time Systems. *IEEE Transactions on Neural Networks*, 19(1):90–106.

Cruz, J. B., Simaan, M. A., Gacic, A., and Liu, Y. (2002). Moving Horizon Nash Strategies for a Military Air Operation. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):989–999.

Davis, M. (1977). *Linear Estimation and Stochastic Control*. Chapman and Hall, London, UK.

Finlayson, B. A. (1972). *The Method of Weighted Residuals and Variational Principles*. Academic Press, New York, NY.

Fleming, W. H. and Rishel, R. W. (1975). *Deterministic and Stochastic Optimal Control*. Springer-Verlag, New York, NY.

Geromel, J. C., de Souza, C. C., and Skelton, R. E. (1998). Static Output Feedback Controllers: Stability and Convexity. *IEEE Transactions on Automatic Control*, 43(1):120–125.

Kailath, T. (1968). An Innovations Approach to Least Square Estimation Part I: Linear Filtering in Additive White Noise. *IEEE Transactions on Automatic Control*, 13(6):646–655.

Klompstra, M. B. (2000). Nash equilibria in risk-sensitive dynamic games. *IEEE Transactions on Automatic Control*, 45(7):1397–1401.

Mukaidani, H., Xu, H., and Dragon, V. (2010). Static Output Feedback Strategy of Stochastic Nash Games for Weakly-Coupled Large-Scale Systems. In *Proc. of the American Control Conference*, pages 361–366, Baltimore, MD.

Sain, M. K. (1966). Control of Linear Systems According to the Minimal Variance Criterion—A New Approach to the Disturbance Problem. *IEEE Transactions on Automatic Control*, AC-11(1):118–122.

Sain, M. K. and Liberty, S. R. (1971). Performance Measure Densities for a Class of LQG Control Systems. *IEEE Transactions on Automatic Control*, AC-16(5):431–439.

Sain, M. K., Won, C.-H., Spencer, Jr., B. F., and Liberty, S. R. (2000). Cumulants and risk-sensitive control: A cost mean and variance theory with application to seismic protection of structures. In Filar, J., Gaitsgory, V., and Mizukami, K., editors, *Advances in Dynamic Games and Applications*, volume 5 of *Annals of the International Society of Dynamic Games*, pages 427–459. Birkhuser Boston.

Sandberg, I. W. (1998). Notes on Uniform Approximation of Time-Varying Systems on Finite Time Intervals. *IEEE Transactions on Circuit and Systems-1:Fundamental Theory and Applications*, AC-45(8):863–865.

Smith, P. J. (1995). A Recursive Formulation of the Old Problem of Obtaining Moments from Cumulants and Vice Versa. *The American Statistician*, (49):217–219.

Van De Water, H. and Willems, J. C. (1981). The Certainty Equivalence Property in Stochastic Control Theory. *IEEE Transactions on Automatic Control*, AC-26(5):1080–1087.

Won, C.-H., Diersing, R. W., and Kang, B. (2010). Statistical Control of Control-Affine Nonlinear Systems with Nonquadratic Cost Function: HJB and Verification Theorems. *Automatica*, 46(10):1636–1645.

Wonham, W. M. (1968). On the Seperation Theorem of Stochastic Control. *SIAM Journal of Control*, 6(2):312–326.

Zheng, D. (1989). Some New Results on Optimal and Suboptimal Regulators of the LQ Problem with Output Feedback. *IEEE Transactions on Automatic Control*, 34(5):557–560.

Zhu, Q., Han, Z., and Basar, T. (2012). A differential game approach to distributed demand side management in smart grid. In *IEEE International Conference on Communications (ICC)*, pages 3345–3350.