ADAPTIVE GAMES

George P. Papavassilopoulos
Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, California 90089-0781

1. INTRODUCTION

The area of adaptive control has received a lot of attention during recent years. Many different schemes have been proposed and studied and several interesting results have been obtained. In almost all the papers the single objective case is addressed: There is one decision maker with his own control objective or there are many controllers acting in a decentralized way who nonetheless have a common objective, i.e., they are a team. Nonetheless, there are cases where there exist many controllers, each one of which has his own objective. Such multiobjective control problems can arise after the decentralization of a large system or exist as such due to the inherent characteristics of the problem. Situations like these belong to the realm of game theory. It is only natural to try to extend the ideas of adaptive control to the area of game theory. As a matter of fact, ignorance of several parameters pertaining to an opponent for which parameters no apriori off line identification is feasible is quite natural in situations of conflict.

There are very few papers in the literature addressing such issues [1,2,4-6]. In the present paper we first introduce a simple example by which we demonstrate several ideas and subsequently we consider some more general situations and describe some results. Section 2 deals with the introductory example.

We consider a scalar linear deterministic evolution equation with two controllers each one being interested in minimizing a one step ahead quadratic cost. The matrix A (here is a scalar) as well as the weight with which each player (i.e., controller) penalizes his control effort are not known to the other player. At each instant of time, each player

This work was supported by ZWO, The Netherlands Organization for the Advancement of Pure Research, Contract Number B62-239, by the U.S. Air Force Office of Scientific Research under Grant AFOSR-85-0254 and by the National Science Foundation under Grant NSF-INT-85-04097.

S. Albeverio et al. (eds.), Stochastic Processes in Physics and Engineering, 223–236.

knows the previous states and his own previous control actions, but not those of his opponent (this is to be contrasted with [1] where the past actions of all the players are known to all of them, i.e., the players have the same information). Each player assumes such a model for the system evolution as if he were the sole decision maker and employs a least squares scheme to estimate the parameter of this model, based on which estimate he calculates his current control action. The question is how the system will behave if the control actions are thus calculated; one essentially has to compare the resulting behavior with that which would result if all the parameters were known to all the players and the Nash concept were employed. Our basic result is that if the closed loop matrix of the known parameter case is asymptotically stable, then the adaptive scheme for the unknown parameter case outlined above with produce an asymptotically stable system, and that the control gains will converge but not to those of the known parameter case. Thus, the closed loop system does not behave as time goes by, like in the known parameter case. This weakness is due to the fact that the standard least squares algorithm pertains really to time invariant systems, whereas the hypothetical system considered by each controller for estimation purposes should be thought of as time varying since the control gains of the ignored player are time varying and are being incorporated in the parameters of the hypothetical system. In Section 3 the more general ARMAX model is considered. The same rationale employed for the example of Section 2 is being utilized here and some results are briefly delineated.

Better or worse results are possible for schemes different than those proposed, but it should be stressed that the purpose of this paper is not as much to provide the best scheme and its complete analysis, but rather to introduce and explain some ideas and demonstrate that positive results are possible for dynamic adaptive games with different information available to the players. Finally, it should be stressed that here we are primarily interested in dynamic cases where the controllers have different information, whereas [1] assumes common information and [2], [3] deal with static cases. Related and more complete results are given in [4-6].

2. INTRODUCTORY EXAMPLE

Consider a system with evoluation equation

$$x_{k+1} = ax_k + u_{1k} + u_{2k}, \quad k = 0,1,2,\ldots, x_0 = \text{given} \tag{1}$$

and two costs

$$J_i = (x_{k+1})^2 + r_i u_{ik}^2, \quad r_i > 0, \quad i = 1,2 \tag{2}$$

All the quantities are scalars. At time k the players know $x_k, x_{k-1}, \ldots, x_0$ and thus the perfect (Nash) solution[1,2] is obtained by solving

[1] The perfect Nash solution for the known parameter case (5) does not change if player i has perfect recall of his past actions.

$$Q \begin{bmatrix} u_{1k} \\ u_{2k} \end{bmatrix} = -a \begin{bmatrix} 1 \\ 1 \end{bmatrix} x_k \tag{3}$$

$$Q = \begin{bmatrix} 1+r_1 & 1 \\ 1 & 1+r_2 \end{bmatrix} \tag{4}$$

Q has determinant $\Delta = r_1 r_2 + r_1 + r_2 \neq 0$ and thus (3) yields

$$\begin{bmatrix} u^*_{1k} \\ u^*_{2k} \end{bmatrix} = -\frac{1}{\Delta} \begin{bmatrix} r_2 \\ r_1 \end{bmatrix} a x_k \tag{5}$$

Substituting u^*_{1k}, u^*_{2k} from (5) into (1) results in the closed-loop system

$$x_{k+1} = a_c x_k \tag{6}$$

$$a_c = a \frac{r_1 r_2}{r_1 r_2 + r_1 + r_2} \tag{7}$$

The closed loop system is asymptotically stable if $|a_c| < 1$. (All the above generalize to the multivariable case with the only exception that the matrix corresponding to Q is not necessarily invertible and thus it has to be assumed to be.) The important thing to notice is that the solution (5) assumes that each player knows x_k, but that also has enough knowledge about a, r_1, r_2. Although it is reasonable for player 1 to know r_1, it is not reasonable to assume that he knows a and even more r_2. We are thus motivated to consider the following situation.

Player 1 knows r_1 and at time k has perfect recall of x_k, x_{k-1}, $\ldots, x_0$, $u_{1,k-1}, \ldots, u_{1,0}$. He assumes that the system obeys

$$x_{k+1} = a_i x_k + u_{ik} , \tag{8}$$

---

[2]The general definition of a Nash equilibrium is as follows: let J_i: $U_1 \times U_2 \to R$. $(u_1^*, u_2^*) \in U_1 \times U_2$ is a Nash equilibrium if $J_1(u_1^*, u_2^*) \leq (u_1, u_2^*)$, $\forall\ u_1^* \in U_1$ and $J_2(u_1^*, u_2^*) \leq J_2(u_1^*, u_2)$, $\forall\ u_2 \in U_2$.

with $i = 1$ in which a_1 is unknown to him. At time k he creates a least squares estimate of a_1, as if a_1 were an unknown constant, see (11), based on which estimate he minimizes J_1 and thus calculates u_{1k}, see (9). Of course, although it is easily seen by player 1 that a_1 is time varying (actually at time k, $a_1 = a - \frac{1}{1+r_2}\hat{a}_k^2$) it is treated for the least squares scheme as a constant in the hope that eventually a_1 converges to a constant. A similar scheme is employed by player 2, see (10), (12) and (8) with $i = 2$. The thus calculated controls act in the real system (1) and therefore the resulting closed loop system is (13). The issue is if and where the sequences described in (9)-(13) converge and the relation of the limits to (5) and (6).

$$u_k^1 = -\frac{1}{1+r_1}\hat{a}_k^1 x_k \tag{9}$$

$$u_k^2 = -\frac{1}{1+r_2}\hat{a}_k^2 x_k \tag{10}$$

$$\hat{a}_{k+1}^1 = \hat{a}_k^1 + \frac{x_k}{\sum_{i=0}^{k} x_i^2}(x_{k+1} - \hat{a}_k^1 x_k - u_k^1), \quad \hat{a}_0^1 = \text{given} \tag{11}$$

$$\hat{a}_{k+1}^2 = \hat{a}_k^2 + \frac{x_k}{\sum_{i=0}^{k} x_i^2}(x_{k+1} - \hat{a}_k^2 x_k - u_k^2), \quad \hat{a}_0^2 = \text{given} \tag{12}$$

$$x_{k+1} = \left(a - \frac{1}{1+r_1}\hat{a}_k^1 - \frac{1}{1+r_2}\hat{a}_k^2\right)x_k, \quad x_0 = \text{given} \tag{13}$$

It should be pointed out that the difference in models assumed by the players — see (8), $i = 1,2$, — is motivated by the information used, by which the players do not know their opponent's past actions; for the same reason the estimates $\hat{a}_k^1$, $\hat{a}_k^2$ are not identical, although both players use the same estimation scheme (contrast with [1]). In some sense, each player employs a "single objective" rationale, when he assumes (8) and minimizes J_i of (2). Let us introduce some notation.

$$\rho_1 = \frac{1}{1+r_1}, \qquad \rho_2 = \frac{1}{1+r_2} \tag{14}$$

$$R = \begin{bmatrix} 1 & \rho_2 \\ \rho_1 & 1 \end{bmatrix}, \quad T = \begin{bmatrix} 1 & -\sqrt{\rho_2/\rho_1} \\ \sqrt{\rho_1/\rho_2} & 1 \end{bmatrix} \tag{15}$$

$$\lambda = 1 + \sqrt{\rho_1\rho_2}\ , \quad \lambda_2 = 1 - \sqrt{\rho_1\rho_2} \tag{16}$$

$$\hat{a}_k = \begin{bmatrix} \hat{a}_k^1 \\ \hat{a}_k^2 \end{bmatrix} \tag{17}$$

$$\beta_k = R\hat{a}_k - a\begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{18}$$

$$\gamma_k = T^{-1}\beta_k \tag{19}$$

$$\sigma_k^2 = \frac{x_k^2}{\sum_{i=0}^{k} x_i^2}\ , \quad \sigma_o^2 = 1 \tag{20}$$

It holds

$$T^{-1}RT = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad 2 > \lambda_1 > 1 > \lambda_2 > 0 \tag{21}$$

Substituting u_k^1, u_k^2 from (9), (10) into (11)-(13) and using (14)-(20) we obtain

$$\gamma_{k+1} = \left(I - \sigma_k^2 \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \right)\gamma_k, \quad \gamma_k = \begin{bmatrix} \gamma_{1k} \\ \gamma_{2k} \end{bmatrix} \tag{22}$$

$$x_{k+1} = \theta_k x_k \tag{23}$$

$$\theta_k = a_c - [\rho_1, \rho_2]\gamma_k \tag{24}$$

Our basic result is the following.

Proposition 1. Let $|a_c| < 1$ (i.e., the closed loop matrix of the known parameter perfect Nash solution is assumed to be asymptotically stable). Then the $\{x_k\}$ sequence generated by (23) converges to zero, i.e., the closed loop system resulting from the adaptive scheme (9)-(13) is also asymptotically stable for any initial conditions x_0, $\hat{a}_0^1$, $\hat{a}_0^2$. Also, the parameters $\hat{a}_k^1$, $\hat{a}_k^2$ converge.

Proof. $0 \le \sigma_k^2 \le 1$. If (i) $\liminf \sigma_k^2 = \sigma^* > 0$, then for $k \ge K(\varepsilon)$: $1 \ge \sigma_k^2 \ge \varepsilon > 0$ for some ε: $0 < \varepsilon < \sigma^*$. This yields: $1 - \lambda_i \le 1 - \lambda_i \sigma_k^2 < 1 - \lambda_i \varepsilon \Leftrightarrow \mp \sqrt{\rho_1\rho_2} \le 1 - (1 \pm \sqrt{\rho_1\rho_2})\sigma_k^2 < 1 - \varepsilon(1 \pm \sqrt{\rho_1\rho_2})$. It holds: $-1 < \mp\sqrt{\rho_1\rho_2}$ and $1 - \varepsilon(1 \pm \sqrt{\rho_1\rho_2}) < 1$ and thus $\gamma_k \to 0$. Consequently, $\theta_k \to a_c$ and since $(\sigma_{k+1})^{-2} = 1 + \sigma_k^{-2}\theta_k^{-2} \cong 1 + \sigma_k^{-2} a_c^{-2}$ with $a_c^{-2} > 1$ we have $\sigma_k^{-2} \to +\infty$ and thus $\sigma_k^2 \to 0$, contradiction. If (ii) $\liminf \sigma_k^2 = 0$ but $\limsup \sigma_k^2 > 0$, then there is a subsequence $\sigma_{n_k}^2 \to \sigma^* > 0$ which implies $|1 - \lambda_i \sigma_{n_k}^{-2}| < 1 - \delta$ for some $\delta < 0$. Since the coefficients $1 - \lambda_i \sigma_{n_k}^2$ appear infinitely often in (22) (argument similar as in (i)), we have $\gamma_k \to 0$. Thus $\sigma_k^2 \to 0$; then $1 - \sigma_k^2 \lambda_i \to 1$ so that eventually $\gamma_{i,k+1}$ and $\gamma_{i,k}$ have the same sign. It always holds: $|\lambda_{i,k+1}| < |\gamma_{i,k}|$ and thus $\gamma_{i,k} \to \gamma_i^*$ monotonically, for some $\gamma^* = (\gamma_1^*, \gamma_2^*)$. Consequently, $\theta_k \to a_c - [\rho_1, \rho_2]\gamma^* = \theta^*$. If $|\theta^*| > 1$, then

$$\sigma_k^2 \cong \frac{(\theta^*)^{2k}}{1 + (\theta^*)^2 + \ldots + (\theta^*)^{2k}} \to 1 - (\theta^*)^{-2}$$

and thus $\sigma_k^2 \to 1 - (\theta^*)^{-2} > 0$, contradiction. Thus $|\theta^*| \le 1$. If $|\theta^*| < 1$ then obviously $x_k \to 0$. If $\theta_k^2 \to 1^+$ then eventually $x_{k+1}^2 = \theta_k^2 x_k^2 \ge x_k^2$ and thus

$$\sigma_k^2 = \frac{1}{\left(\frac{x_o}{x_k}\right)^2 + \ldots + \left(\frac{x_{k-1}}{x_k}\right)^2 + 1} \ge \frac{1}{1 + \ldots + 1 + 1} = \frac{1}{k+1}$$

It then holds: $0 \le 1 - \lambda_i \sigma_k^2 \le 1 - \frac{\lambda_i}{k+1} \Rightarrow |\gamma_{i,k+1}| \le \left(1 - \frac{\lambda_i}{k+1}\right)|\gamma_{i,k}| \Rightarrow \gamma_{ik} \to 0 \Rightarrow \gamma_1^* = \gamma_2^* = 0 \Rightarrow \theta^* = a_c$. But $|a_c| < 1$ and this contradicts $|\theta^*| = 1$.

So θ_k^2 cannot converge to 1^+. If $\theta_k^2 \to 1^-$, then eventually, $x_{k+1}^2 \le x_k^2$ and thus x_k^2 converges to some $(x^*)^2$. If $x^* \neq 0$ then $\sigma_k^2 = \frac{x_k^2}{x_o^2 + \ldots + x_k^2} \ge \frac{(x^*)^2}{x_o^2(k+1)}$ and an argument similar to the one above yields $\gamma^* = 0$, $\theta^* = a_c$ and thus $x^* = 0$, contradiction. Thus if $\theta_k^2 \to 1^-$ then $x_k \to 0$. Finally, consider the case $\theta_k^2 \to 1$ but not from above or below. In particular, let $\theta_k \to -1 = \theta^*$. It holds

$$\theta_{k+1} = \theta_k + \sigma_k^2[\rho_1\lambda_1, \rho_2\lambda_2]\gamma_k \tag{25}$$

Since θ_k does not converge to -1 from above or below it must be

$$[\rho_1\lambda_1, \rho_2\lambda_2]\gamma^* = 0 \tag{26}$$

$\theta^* = -1$ means

$$a_c - [\rho_1, \rho_2]\gamma^* = 0 \tag{27}$$

(26) and (27) can be solved for γ_1^*, γ_2^* to yield

$$\gamma^* = \frac{-1 + a_c}{2\rho_1\rho_2\sqrt{\rho_1\rho_2}} \begin{bmatrix} -\rho_2\lambda_2 \\ \rho_1\lambda_1 \end{bmatrix}, \quad \gamma_1^* > 0, \quad \gamma_2^* < 0 \tag{28}$$

From (22) we obtain

$$\frac{\gamma_{1,k+1} - \gamma_{1,k}}{\gamma_{2,k+1} - \gamma_{2,k}} = \frac{\lambda_1\gamma_{1k}}{\lambda_2\gamma_{2k}} \tag{29}$$

Since θ_k does not converge to -1 from above or below for infinitely many k's it holds

$$\theta_{k+2} < \theta_{k+1}, \qquad \theta_{k+1} > \theta_k \tag{30}$$

which using (25) implies

$$\lambda_1\rho_1\gamma_{1,k+1} + \lambda_2\rho_2\gamma_{2,k+1} < 0 \tag{31}$$

$$\lambda_1\rho_1\gamma_{1k} + \lambda_2\rho_2\gamma_{2k} > 0 \tag{32}$$

Since $\gamma_1^* > 0$, $\gamma_2^* < 0$, for k sufficiently large, it holds $\gamma_{1k} > 0$, $\gamma_{2k} < 0$ and thus (31), (32) yields

$$0 > \frac{\gamma_{2k}}{\gamma_{1k}} > \frac{-\lambda_1\rho_1}{\lambda_2\rho_2} > \frac{\gamma_{2,k+1}}{\gamma_{1,k+1}} \tag{33}$$

(29) yields

$$\frac{\gamma_{2,k+1}}{\gamma_{1,k+1}} = \frac{\lambda_2}{\lambda_1}\frac{\gamma_{2k}}{\gamma_{1k}} - \frac{\lambda_2 - \lambda_1}{\lambda_1}\frac{\gamma_{1k}}{\gamma_{1,k+1}} \tag{34}$$

Inserting $\gamma_{2,k+1}/\gamma_{1,k+1}$ from (34) into (33) yields

$$0 > \frac{\gamma_{2k}}{\gamma_{1k}} > -\frac{\lambda_1\rho_1}{\lambda_2\rho_2} > \frac{\lambda_2}{\lambda_1}\frac{\gamma_{2k}}{\gamma_{1k}} + \frac{\lambda_1 - \lambda_2}{\lambda_1}\frac{\gamma_{1k}}{\gamma_{1,k+1}} \tag{35}$$

from which

$$\frac{\gamma_{2k}}{\gamma_{1k}}\frac{\lambda_1 - \lambda_2}{\lambda_2} > \frac{\lambda_1 - \lambda_2}{\lambda_2}\frac{\gamma_{1,k}}{\gamma_{1,k+1}}$$

or (since $\lambda_1 > \lambda_2$)

$$\frac{\gamma_{2k}}{\gamma_{1k}} > \frac{\gamma_{1k}}{\gamma_{1,k+1}}$$

which means that a negative number is greater than a positive one, contradiction. Thus if $\theta_k \to -1$ it must be $\theta_k^2 \to 1^-$. Similarly, we conclude for the case where $\theta_k \to +1$, that $\theta_k^2 \to 1^-$.

In conclusion, we have shown that $\sigma_k^2 \to 0$, $\gamma_k \to \gamma^*$, $x_k \to 0$, $\theta_k \to \theta^*$, $\theta^* \in [-1,1]$ and if $\theta^* = \pm 1$ then eventually $\theta_k \in (-1,1)$. Finally, let us notice that the convergence of γ_k implies that of $\hat{a}_k^1$, $\hat{a}_k^2$, see (17)-(19). □

Remark 1. Ideally, we would like $\gamma_k \to \gamma^* = (0,0)'$, since then the control gains and the closed loop matrix of the adaptive scheme converge to those of the known parameter case. (This is easily verified by solving (18) with $\beta_k = (0,0)$, for $\hat{a}_k^1$, $\hat{a}_k^2$, using (9), (10) and comparing with (5).) This is not achieved by the scheme described here, as the following argument demonstrates (see also Examples in Remark 3): Let γ_{10}, γ_{20} be sufficiently small. Since $|\gamma_{1k}|$, $|\gamma_{2k}|$ are decreasing and converge, $|\theta^*| < 1$. Then, $\sigma_k^2 \cong (\theta^*)^{2k}/(1 + \ldots + \theta^{*2k}) \Rightarrow \sigma_k^2 \cong (1 - \theta^{*2})\theta^{*2k}$ and thus $\gamma_{i,k+1} \cong \gamma_{i,k}[1 - \lambda_i(1 - \theta^{*2})\theta^{*2k}]$. Let $\omega = \theta^{*2}$, $\beta = \lambda_i(1 - \theta^{*2})$. We know that the sequence $y_{n+1} = (1 - \beta\omega^n)y_n$, $\beta > 0$, $0 < \omega < 1$ does not go to zero except if $1 - \beta\omega^n = 0$ for some n. Thus, $\gamma_{i,k} \to \gamma_i \neq 0$ and thus $\theta^* \neq a_c$. Choosing γ_{io} small is obviously desirable, but this means β_k small, i.e., $\hat{a}_k$ close to $a R^{-1}\begin{bmatrix}1\\1\end{bmatrix}$, which is exactly what the players who choose $\hat{a}_0^1$, $\hat{a}_0^2$ do not know.

Remark 2. In our example we considered that the target trajectories are zero (see (2)). One can consider though $J_i = (x_{k+1} - x^*_{i,k+1})^2 + v_i u_{ik}^2$ instead of (2). Then (8) could be modified to $x_{k+1} = a_i x_k + u_{ik} + c_i$ where both a_i and c_i are to be estimated.

Remark 3, Examples. We conducted several examples which demonstrate the convergence properties of (20), (22), (23), (24).

i) For $(a_c, \gamma_{10}, \gamma_{20}, \rho_1, \rho_2) = (0.3, -1, 5, 0.5, 0.3)$, $(\sigma_k^2, \gamma_{1k}, \gamma_{2k}, \theta_k)$ converged to (0, 0.1891, 1.4762, -0.2374) in 6 iterations.

ii) For $(a_c, \gamma_{10}, \gamma_{20}, \rho_1, \rho_2) = (0.3, 3, 6, 0.5, 0.3)$, $(\sigma_k^2, \gamma_{1k}, \gamma_{2k}, \theta_k)$ converged to (0, 0.2767, 1.0231, -0.1453) in 5 ierations.

iii) For $(a_c, \gamma_{10}, \gamma_{20}, \rho_1, \rho_2) = (0.3, 3, 6, 0.5, 0.5)$, $(\sigma_k^2, \gamma_{1k}, \gamma_{2k}, \theta_k)$ converged to (0, 0.3733, 1.3430, -0.5582) in 12 iterations.

iv) For $(a_c, \gamma_{10}, \gamma_{20}, \rho_1, \rho_2) = (0.9, 3, 6, 0.5, 0.3)$, $(\sigma_k^2, \gamma_{1k}, \gamma_{2k}, \theta_k)$ converged to (0, 0.0954, 0.8223, 0.6056) in 12 iterations.

v) For $(a_c, \gamma_{10}, \gamma_{20}, \rho_1, \rho_2) = (-0.9, 3, 6, 0.5, 0.3)$ $(\sigma_k^2, \gamma_{1k}, \gamma_{2k}, \theta_k)$ converged to $(0, 0.0011, 0.1176, -0.9358)$ in 76 iterations.

It should be noticed that θ_k does not converge to a_c (see Remark 1 also). It seems that a_c close to ± 1 slows down the convergence as a comparison of (ii) to (iv) and (v) indicates.

3. THE ARMAX CASE

In this section we discuss the generalization of the introductory example to the ARMAX case. Consider the evolution equation

$$\begin{aligned} y_{t+1} &= a_o y_t + a_1 y_{t-1} + \ldots + a_n y_{t-n} + u_{1t} + u_{2t} + v_t \\ &= A(q^{-1}) y_t + u_1 + u_{2t} + v_t \end{aligned} \tag{36}$$

and the two costs

$$J_i = E[y_{t+1}^2 + r_i u_{it}^2], \quad i = 1,2 \tag{37}$$

where $\{v_t\}$ is a sequence of independent gaussian random variables. If the parameter vector $\theta = (a_o, a_1, \ldots, a_n, r_1, r_2)'$ is known to both players and at each instant of time they know all the previous history $\{y_t, y_{t-1}, \ldots\}$, the Nash equilibrium is

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = - \begin{bmatrix} 1+r_1 & 1 \\ 1 & 1+r_2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} A(q^{-1}) y_t \tag{38}$$

and the resulting closed loop system is

$$y_{t+1} = \left(1 - [1,1] \begin{bmatrix} 1+r_1 & 1 \\ 1 & 1+r_2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) A(q^{-1}) y_t + v_t \tag{39}$$

(We assume that $r_1 r_2 + r_1 + r_2 > 0$ and that the closed loop system matrix of (39) is asymptotically stable.) Let us now consider that player 1 knows r_1, but not $a_o, \ldots, a_n, r_2$, so that he cannot find his control action u_{1t} at time t by using (38). We also assume that at time t he knows $y_t, y_{t-1}, \ldots$ and $u_{1,t-1}, u_{1,t-2}, \ldots$, but not $u_{2,t-1}, u_{2,t-2}, \ldots$.

We follow the same line of development as in the introductory example: player 1 assumes that the system obeys

$$y_{t+1} = a_o^1 y_t + a_1^1 y_{t-1} + \ldots + a_n^1 y_{t-n} + u_{1t} + v_t^1 \tag{40}$$

This is a hypothetical system. If (40) were the true system and player's 1 cost were (37) with $i = 1$, he would use at time t the control action:

$$-\frac{1}{1+r_1}(a_o^1 y_t + \ldots + a_n^1 y_{t-n}) \tag{41}$$

Using the true outputs of the system $y_t, y_{t-1}, \ldots$ and his own previous actions $u_{1,t-1}, \ldots$ he creates estimates — according to some scheme — of $a_o^1, \ldots, a_n^1$, call them $\hat{a}_o^{1t}, \ldots, \hat{a}_n^{1t}$ and he uses them in (41), so that he employs the control action:

$$u_{1t} = \bar{u}_{1t} \overset{\text{det}}{=} -\frac{1}{1+r_1}(\hat{a}_o^{1t} y_t + \ldots + \hat{a}_n^{1t} y_{t-n}) \tag{42}$$

Similarly, player 2 uses

$$u_{2t} = \bar{u}_{2t} \overset{\text{det}}{=} -\frac{1}{1+r_2}(\hat{a}_o^{2t}, y_t + \ldots + \hat{a}_n^{2t} y_{t-n}) \tag{43}$$

Both u_{1t}, u_{2t} of (42), (43) are applied to the real system, so that we have

$$y_{t+1} = a_o y_t + a_1 y_{t-1} + \ldots + a_n y_{t-n} + \bar{u}_{1t} + \bar{u}_{2t} + v_t \tag{44}$$

We also have

$$(\hat{a}_o^{1t}, \ldots, \hat{a}_n^{1t}) = F_1(\hat{a}_o^{1,t-1}, \hat{a}_1^{1,t-1}, \ldots, \hat{a}_n^{1,t-1}, u_{1,t-1}, u_{1,t-2}, \ldots, y_t, y_{t-1}, \ldots) \tag{45}$$

$$(\hat{a}_o^{2t}, \ldots, \hat{a}_n^{2t}) = F_2(\hat{a}_o^{2,t-1}, \ldots, \hat{a}_n^{2,t-1}, u_{2,t-1}, u_{2,t-2}, \ldots, y_t, y_{t-1}, \ldots) \tag{46}$$

where F_1, F_2 are determined by the estimation schemes employed by the two players. The question now is how the behavior of (42)-(46) is related to the behavior of (38), (39). The basic result, for the proof of which we refer to [5,6], is that if the closed loop system of the known parameter case (39) is asymptotically stable, then the control actions (42), (43) and the system (44) behave in the limit like (38), (39), if the estimation schemes F_1, F_2 of (45), (46) pertain to time varying systems (such as weighted least squares, or stochastic approximation types). The important thing to notice is that the hypothetical system (40) should not be thought of as time invariant, since the $a_o^1,..,a_n^1$ incorporate in them not only the $a_o,..a_n$ but also the time varying gains of player 2, i.e., $a_k^1 \cong a_k - \frac{1}{1+r_2} a_k^{2,t}$.

A more general case is to consider

$$\begin{aligned} y_{t+1} &= a_o y_t + ..+ a_n y_{t-n} + b_{10} u_{1t} + \ldots + b_{1,k_1} u_{1,t-k_1} \\ &+ b_{20} u_{2t} + \ldots + b_{2,k_2} u_{2,t-k_2} + v_t \\ &= A(q^{-1}) y_t + B_1(q^{-1}) u_{1k} + B_2(q^{-1}) u_{2t} + v_t \end{aligned} \tag{47}$$

instead of (36), and the costs are still as in (37). Player 1 knows r_1, but ignores $(a_o,..,a_n, b_{10},..,b_{1,k_1}, b_{20},..,b_{2,k_2})$ and r_2, but he knows the basic structure of (47) and thus n, k_1, k_2. At time t, player 1 knows $y_t, y_{t-1},\ldots$ and $u_{1,t-1}, u_{1,t-2},\ldots$, but ignores the previous actions $u_{2,t-1}, u_{2,t-2},..$ of player 2. (Similarly for player 2.) Notice, that lack of knowledge of the other's previous actions is even more pertinent to the model (47) which has delays in the controls, than to (36). Nonetheless, the same rationale can be employed as before, but instead of the hypothetical system (40) which has the same number of delays in y_t as (36), player 1 considers the hypothetical system

$$\begin{aligned} y_{t+1} &= a_o^1 y_t + ..+ a_{\ell_1}^1 y_{t-\ell_1} + b_{10}^1 u_{1,t} + b_{1,1}^1 u_{1,t-1} + \ldots \\ &+ b_{1,m_1}^1 u_{1,t-m_1} + c_o v_t^1 + c_1 v_{t-1}^1 + \ldots + c_{1,t-p_1} v_{t-p_1}^1 \end{aligned} \tag{48}$$

where ℓ_1 does not necessarily equal n. The reason is the following — if all the parameters and previous control actions are known to both the players, the Nash equilibrium is given by solving for u_{1t}, u_{2t}:

$$(A(q^{-1}) y_t + B_1(q^{-1}) u_{1t} + B_2(q^{-1}) u_{2t}) b_{10} + r_1 u_{1t} = 0 \tag{49}$$

$$(A(q^{-1}) y_t + B_1(q^{-1}) u_{1t} + B_2(q^{-1}) u_{2t}) b_{20} + r_2 u_{2t} = 0 \tag{50}$$

from which we obtain:

$$u_{2t} = \frac{r_1 b_{20}}{r_1 r_2 + r_1 b_{20} B_2(q^{-1}) + r_2 b_{10} B_1(q^{-1})} A(q^{-1}) y_t$$

which, when substituted in (47), results to a system of the form (48). This substitution determines ℓ_1, m_1, p_1. Simulation studies based on the above rationale show that the scheme works well, under some assumptions such as asymptotic stability of the closed loop system of the known parameter case and some weak coupling conditions on the coefficients of B_1, B_2, but no rigorous theoretical analysis is currently available, see [6].

Remark. Cases where y_t is vector valued and the costs (37) are substituted by $E[(y_{t+1} - y^{*i}_{t+1})^T Q_i (y_{t+1} - y^{*i}_{t+1}) + r_i \|u_{it}\|^2]$, i.e., where the players have different target objectives — $\{y^{*i}_t\}$ — have been considered in [6].

4. CONCLUSIONS

The aim of this paper was to introduce some ideas pertaining to adaptive schemes for multiobjective control problems with information decentralization. Some basic models and partial results toward this end were presented. Further directions of research may involve the examination of different rationales for creating adaptive schemes, the completion of the theoretical analysis of some of the schemes presented, etc.

5. REFERENCES

[1] Y. M. Chan, "Self Tuning Methods for Multiple Controller Systems," Ph.D. Thesis, University of Illinois at Urbana-Champaign, Dept. of Electrical Engineering, 1981.

[2] T. L. Ting, J. B. Cruz, Jr. and R. A. Milito, "Adaptive Incentive Controls for Stackelberg Games with Unknown Cost Functionals," American Control Conference, San Diego, CA, June 1984.

[3] G. P. Papavassilopoulos, "Iterative Techniques for the Nash Solution in Quadratic Games with Unknown Parameters," accepted to appear in SIAM Journal on Optimization and Control, 1985; also, presented at the Istanbul Workshop on Large Scale Systems, Istanbul, Turkey, June 1984.

[4] G. P. Papavassilopoulos, "Adaptive Dynamic Nash Games: An Example," Seventh Annual Meeting of the Society for Economic Dynamics and Control, London, June 1985.

[5] W. Y. Yang and G. P. Papavassilopoulos, "Decentralized Adaptive Control in a Game Situation for Discrete Time Linear, Time In-

variant Systems," submitted for publication, 1986.

[6] W. Y. Yang, "Decentralized Adaptive Control in a Game Situation," Ph.D. Thesis, University of Southern California, Dept. of EE-Systems, Los Angeles, CA, August 1986.