EQUILIBRIUM CONCEPTS IN STOCHASTIC MULTICRITERIA OPTIMIZATION

DAVID A. CASTANON
Electronic Systems Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

This paper explores some issues in equilibrium solutions of stochastic multicriteria decision problems. Correct treatment of this problem involves an extended notion of state, thereby altering conventional concepts of open-loop and closed-loop decisions. Restrictions on transfer of information lead to signalling-free concepts. These topics are discussed and illustrated with a simple example.

1. INTRODUCTION

A common approach in the study of decentralized decision-making in optimization problems is to model the individual decision makers as players within a game [2]. Such an approach is appropriate for classes of systems where there are multiple criteria for multiple decision makers, decentralized information, and natural hierarchies in the order of decision-making. Recent works in the field [1], [5], [6], have focused on systems with a sequential decision structure; these decisions are commonly referred to as "Stackelberg" solutions. Most of these works have dealt with deterministic systems, or with systems where the stochastic structure is very simple. This paper discusses solutions of systems with sequential decision structure and nontrivial stochastic information structure. In particular, it defines and examines three classes of equilibrium solutions, providing a simple example which highlights their differences.

2. DEFINITIONS

This section formally defines some basic terms, and formulates the sequential decision problem.

Definition 2.1. An n-player game is a triple G: $\{\underline{D}, J, SR\}$, where $\underline{D} = D_1 \times D_2 \times \ldots \times D_N$ is a decision space, D_i is the set of decisions for player i, J is a vector, extended-real valued performance criteria.

$$J: D \to (R^e)^N$$

J_i is the criterion of the i^{th} player. SR is a selection rule which selects a subset $S^* \subset \underline{D}$ as a solution of the game.

The essential problem of solving a game consists of obtaining S*. The selection rule (also known as solution concept) is usually based on behavioral principles governing the players in the game; examples of these principles are Nash, Pareto, minimax, and Stackelberg principles.

Definition 2.2. An Equilibrium selection rule is a selection rule such that

$$\underline{d}^* = (d_1^*, \ldots d_N^*) \in S^*$$

if $J_i(\underline{d}^*) \leq J_i(\underline{d}_i)$ for all $\underline{d}_i \in \underline{D}$ of the form

$$\underline{d}_i = (d_1^*, \ldots, d_{i-1}^*, d_i, d_{i+1}^*, \ldots d_N^*) \quad \text{for all } i = 1, \ldots N.$$

The preparation of this paper was supported in part by ONR Contract N00014-76-C-0346.

Equilibrium solutions are thus defined in terms of performance criteria and admissible strategies. Consider now the following stochastic sequential dynamical system:

(A) State equation

$$x_{t+1} = f_t(x_t,\ u_t^1,\ldots,u_t^n,\ \theta_t) \tag{2.1}$$

Observation equation

$$y_t^i = g_t^i(x_t,\ \xi_t^i) \tag{2.2}$$

Loss functionals

$$L_0^i = \sum_{t=0}^{T-1} h_t^i(x_t,\ u_t^1,\ldots,u_t^n) + h_T^i(x_T) \tag{2.3}$$

The variables x_t, y_t^i, u_t^i are elements of appropriately defined state, observation and decision spaces, assumed to be finite-dimensional Euclidean spaces. The variables θ_t, ξ_t^i and x_0 are random variables, with the functions f_t, g_t^i and h_t^i assumed appropriately measurable. The active information available to each decision-maker i at time t is denoted I_t^i, where

$$I_0^i = \{y_0^i,\ u_0^1,\ldots u_0^{i-1}\} \tag{2.4}$$

$$I_t^i = I_{t-1}^i \cup \{y_t^i,\ u_{t-1}^{i+1},\ldots u_{t-1}^n,\ u_t^0,\ldots u_t^{i-1}\} \tag{2.5}$$

Each player knows the value of the previous decisions made by all players. At each stage, decision are made and announced sequentially. This sequential information pattern is typical of Stackelberg dynamic games [3], [4]. The stochastic sequential dynamical system described in equations (2.1) to (2.5) is a typical model of hierarchical system [6]. However, this mathematical description does not constitute a game; performance criteria, admissible strategies and selection rules must be specified.

<u>Definition 2.3</u>. An open-loop strategy λ_t^i is a measurable map from the space of information sets to the space of decision such that

$$u_t^i = \lambda_t^i(I_t^i).$$

<u>Definition 2.4</u>. An open-loop equilibrium solution to the dynamic decision model (A) is a member of the equilibrium solution set of the game G_{OL}, where

$$G_{OL} := \{\underline{D}_{OL},\ \underline{J},\ SR_{eq}\}$$

and

$$D_{OL}^i = \{\text{open-loop } \underline{\lambda}^i = (\lambda_0^i,\ldots\lambda_{T-1}^i)\}$$

$$\underline{\lambda} = (\underline{\lambda}^1,\ldots\underline{\lambda}^n)$$

$$J^i(\underline{\lambda}) = E^{\underline{\lambda}}(L^i)$$

and $E^{\underline{\lambda}}$ is the expectation with respect to the probability distribution induced by the strategy $\underline{\lambda}$.

Notice that, even through the admissible strategies are functions of available information, the solution is termed open-loop. This distinction is made because the information sets I_t^i do not truly represent the state of the system; they do not include knowledge of the previous strategies used in the game by other players.

Definition 2.5: A closed-loop strategy γ_t^i is a strategy-dependent map which maps $\{\underline{\gamma}_t^i\}$ and I_t^i into an admissible decision u_t^i, where

$$\underline{\gamma}_t^i = \{\gamma_0^1, \ldots \gamma_0^n, \gamma_1^1, \ldots \gamma_{t-1}^n, \gamma_t^1, \ldots \gamma_t^{i-1}\}$$

$$u_t^i = \gamma_t^i(I_t^i; \underline{\gamma}_t^i)$$

The key difference between open- and closed-loop strategies is the dependence of closed-loop strategies on previous strategies, because the true state of the system is a probability distribution (conditional or unconditional) depending on previous strategies.

Definition 2.6: A closed-loop equilibrium solution to the dynamic decision model (A) is a member of the equilibrium solution set of the game G_{CL}, where

$$G_{CL}: = \{\underline{D}_{CL}, \underline{J}_{CL}, SR_{eq}\}$$

and $D_{OL}^i: = \{\text{closed-loop } \underline{\gamma}^i = (\gamma_0^i, \ldots \gamma_{T-1}^i)\}$

$$\underline{\gamma} = (\underline{\gamma}^1, \ldots \underline{\gamma}^n)$$

$$J^i(\underline{\gamma}) = E^{\underline{\gamma}}(L^i)$$

$E^{\underline{\gamma}}$ is the expectation with regard to the induced probability distribution.

Occasionally, the dependence of closed-loop strategies on previous strategies may be summarized through a transformation to a standard form, as in Witsenhausen [7], or Castanon and Sandell [1]. The original stochastic model is transformed to an equivalent deterministic model, defining the unconditional probability distribution of the state, x_t, denoted by π_t, as the true state.

The closed-loop strategies are represented as

$$u_t^i = \gamma_t^i(I_t^i; \pi_t)$$

The equilibrium solutions obtained with this representation of admissible strategies are included in the equilibrium solutions obtained in definition (2.6); the converse is not true in general.

This approach requires a non-trivial transformation of the system dynamics to a standard form, which is not always possible to do. An alternative approach is to use another representation of closed-loop strategies, making γ_t^i dependent on the conditional distribution of the state, given information I_t^i and previous strategies $\underline{\gamma}_t^i$. That is,

$$u_t^i = \gamma_t^i(\hat{\pi}_t^i, I_t^i)$$

where

$$\hat{\pi}_t^i = \text{probability distribution of } x_t, \text{ given } I_t^i \text{ and } \underline{\gamma}_t^i$$

The form of closed-loop strategies implies that the players have exact knowledge of all strategies used by other players previously. This enables the players to extract information about the system state from observation of the previous decisions. This is known as signalling [4]. In many situations, it is impractical or inaccurate to use strategies which require exact knowledge of past strategies; furthermore, it may be desirable for players to choose their strategies on the basis of internal perceptions of cost (delayed commitment) rather than an external cost such as the unconditional expectation of the loss function.

Definition 2.7: A signalling-free conditional probability distribution for player i at stage t, denoted p_t^i, is a version of the conditional probability distribution of the system variables assuming all $u_k^j \in I_t^i$ were produced by constant strategies.

Definition 2.8: A signalling-free equilibrium solution to the dynamic decision model (A) is a member of the equilibrium solution set of the game G_{SF}, where

$$G_{SF}: = \{\underline{D}_{SF}, \underline{J}_{SF}, SR_{eq}\}$$

$$D_{SF}^i: = \{\gamma_t^i(I_t^i; p_t^i)\}$$

$$\underline{\gamma}_t^i = \{\gamma_t^i, \ldots \gamma_t^n, \gamma_{t+1}^1, \ldots \gamma_T^n\}$$

$$J_t^i(\underline{\gamma}_t^i) = E^{p_t^i, \underline{\gamma}_t^i}(L^i)$$

where $E^{p_t^i, \underline{\gamma}_t^i}$ is the expectation with respect to the conditional probability distribution p_t^i, assuming future decisions are defined using strategies $\underline{\gamma}_t^i$.

A key point to note is that signalling-free equilibrium solutions are equilibria between nT players, reflecting the fact that decisions are made stagewise based on internal perceptions. Also, the equilibrium properties of the strategies are not lost if at any prior stage, non-equilibrium strategeis were used. This is possible because of the lack of signalling; the solution (and the relevant probability densities) can be obtained recursively dealing exclusively with the information sets I_t^i and future strategies $\underline{\gamma}_t^i$.

3. ILLUSTRATIVE EXAMPLES

Consider the discrete-time system described by the state equations

$$x_1 = x_0 + u_0 + v_0 + \theta_0 \quad (3.1)$$

$$x_2 = x_1 + u_1 + \theta_1 \quad (3.2)$$

with observations

$$z_0 = x_0 + \xi_0 \quad (3.3)$$

$$z_1 = x_1 \quad (3.4)$$

where x_0, θ_0, θ_1, ξ_0 are independent zero-mean Gaussian random variables with covariances Σ_0, Θ_0, Θ_1, Ξ_0 respectively, and all variables are real-valued scalars.

The players' active information sets are

$$I_0^1 = \{z_0\} \quad (3.5)$$

$$I_1^1 = \{x_1, z_0, v_0\} \quad (3.6)$$

$$I_0^2 = \{u_0\} \quad (3.7)$$

where player u is denoted 1, player v is denoted 2. The loss functionals are

$$L^1 = x_2^2 + u_0^2 + u_1^2 \quad (3.8)$$

$$L^2 = x_2^2 + v_0^2 \quad (3.9)$$

The linear quadratic Gaussian nature of this example implies that there are open-loop and closed-loop stochastic equilibria of the form

$$u_0 = az_0 \tag{3.10}$$

$$v_0 = bu_0 \tag{3.11}$$

$$u_1 = cx_1 \tag{3.12}$$

Additionally, the signalling-free equilibria will also be in that form. Appendix A computes the signalling-free equilibrium strategies and the associated unconditional losses. Appendices B and C compute the linear closed-loop and open-loop strategy and the unconditional losses, respectively. The results are summarized in Table 3.1 when $\Xi_0 = \Sigma_0 = \Theta_0 = \Theta_1 = 1$.

Table 3.1

	Signalling-free Equilibrium	Closed-loop Equilibrium	Open-loop Equilibrium
a*	$-\frac{5}{33}$	$-\frac{4}{33}$	$-\frac{2}{33}(\frac{3}{2} \pm\sqrt{3})$
b*	$-\frac{1}{5}$	$\frac{5}{8}$	$\frac{-1 \pm\sqrt{3}}{2}$
c*	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
J^1	1.94	1.87	1.86 or 2.03
J^2	1.45	1.43	1.46 or 1.51

4. CONCLUSION

The concepts of open-loop and closed-loop strategies in sequential stochastic games were adapted from deterministic games with one principle in mind: Closed-loop strategies are maps which depend on the state of the system. In stochastic systems, this state is often a probability distribution. These concepts also apply to non-sequential games. The example in Section 3 and Appendices B and C highlight the differences between open-loop and closed-loop equilibrium strategies. In particular, open-loop equilibria are obtained using variational principles; closed-loop equilibria can be obtained through stagewise decomposition similar to dynamic programming.

A third class of equilibrium solutions was defined, based on the concept of "delayed commitment" and no signalling. This solution is easier to compute in general because of the restrictions on signalling and the recursive nature of the definition.

There are many open areas of research. Castanon [4] has established an equivalence between signalling-free and closed-loop equilibria for special information patterns. It is not clear that open-loop equilibria are also closed-loop equilibria in stochastic games, as is true in deterministic games [8]. The dependence of closed-loop equilibria on the representation of the true system state must be studied in depth. Understanding these questions is essential in formulating a unified theory of generalized decision-making.

APPENDIX A

Consider the problem of obtaining a signalling-free equilibrium for the system described in Section 3.

$$J_1^1 = E\{u_1^2 + x_2^2 | x_1\} = u_1^2 + (x_1 + u_1)^2 + \Theta_1 \quad \text{(A.1)}$$

Hence,

$$u_1^* = -\frac{1}{2} x_1 \quad \text{(A.2)}$$

Similarly,

$$J_0^2 = E\{v_0^2 + x_2^2 | u_0 = \text{constant}, u_1 = -\frac{1}{2} x_1\}$$

$$= \Theta_0 + \Theta_1 + \frac{1}{4}(u_0 + v_0)^2 + v_0^2 + \Sigma_0 \quad \text{(A.3)}$$

because x_0 is zero-mean. Thus,

$$v_0^* = -\frac{1}{5} u_0 \quad \text{(A.4)}$$

For player u at stage 0, one gets

$$J_0^1 = E\{x_2^2 + u_1^2 + u_0^2 | z_0, v_0 = -\frac{1}{5} u_0, u_1 = -\frac{1}{2} x_1\}$$

$$= \frac{1}{2} \bar{x}_1^2 + u_0^2 + \Sigma_1^0 \quad \text{(A.5)}$$

where

$$\hat{x}_0 = \frac{\Sigma_0}{\Sigma_0 + \Xi_0} z_0 \quad \text{(A.6)}$$

$$\bar{x}_1 = \hat{x}_0 + \frac{4u_0}{5} \quad \text{(A.7)}$$

$$\Sigma_1^0 = \frac{\Sigma_0 \Xi_0}{\Sigma_0 + \Xi_0} + \Theta_0 \quad \text{(A.8)}$$

Hence,

$$u_0^* = \frac{-10}{33} \hat{x}_0 = \frac{-10}{33} \frac{\Sigma_0}{\Sigma_0 + \Xi_0} z_0 \quad \text{(A.9)}$$

Assuming $\Sigma_0 = \Xi_0 = \Theta_0 = \Theta_1 = 1$, the unconditional costs can be computed as:

$$E\{u_0^2\} = \frac{1}{2} \cdot \frac{10}{33}^2 \quad \text{(A.10)}$$

$$E\{v_0^2\} = \frac{1}{2} \cdot \frac{2}{33}^2 \quad \text{(A.11)}$$

$$E\{x_1^2\} = 2 - \frac{8 \cdot 29}{(33)^2} \quad \text{(A.12)}$$

Thus,

$$J_1 = 2 + \frac{1}{2}(\frac{10}{33})^2 - \frac{4 \cdot 29}{(33)^2} = 2 - \frac{2}{33} \quad \text{(A.13)}$$

$$J_2 = \frac{3}{2} - \frac{58}{(33)^2} + \frac{2}{(33)^2} = \frac{3}{2} - \frac{56}{(33)^2} \quad \text{(A.14)}$$

APPENDIX B

Consider the problem of obtaining a closed-loop equilibrium solution for the model in Section 3, using strategies with conditional probability distribution dependence.

$$J_1^1 = E\{u_1^2+x_2^2 | a, b, z_0, x_1\} = u_1^2 + (x_1+u_1)^2 + \Theta_1 \tag{B.1}$$

Thus,
$$u_1^* = -\frac{1}{2} x_1 \tag{B.2}$$

$$J_0^2 = E\{x_2^2 + u_0^2 | az_0 = u_0, u_1 = -\frac{1}{2} x_1\} \tag{B.3}$$

If a=0, then b=0. Thus, assume $a \neq 0$. In this case, knowledge of u_0 implies knowledge of z_0. Thus

$$J_0^2 = \frac{\bar{x}_1^2}{4} + v_0^2 + \Theta_1 + \Sigma_1^0 \tag{B.4}$$

where

$$\hat{x}_0 = \frac{\Sigma_0}{\Sigma_0+\Xi_0} z_0 \tag{B.5}$$

$$\bar{x}_1 = \hat{x}_0 + u_0 + bu_0 \tag{B.6}$$

$$\Sigma_1^0 = \frac{\Sigma_0\Xi_0}{\Sigma_0+\Xi_0} + \Theta_0 \tag{B.7}$$

Thus, $v_0^* = -\frac{1}{5}(\hat{x}_0 + az_0) = -\frac{1}{5}(az_0)(1 + \frac{\Sigma_0}{a(\Sigma_0+\Xi_0)})$ (B.8)

Now,

$$J_0^1 = E\{x_2^2+u_0^2+u_1^2 | z_0, u_1 = -\frac{1}{2}x_1, v_0 = -\frac{1}{5} u_0(b^*)\}$$
$$= \frac{\bar{x}_1^2}{2} + u_0^2 + \Sigma_1^0 \tag{B.9}$$

where
$$\bar{x}_1 = \frac{4}{5}(u_0 + \hat{x}_0) \tag{B.10}$$

so,
$$u_0^* = -\frac{8}{33} z_0 \frac{\Sigma_0}{\Sigma_0+\Xi_0} \tag{B.11}$$

When $\Xi_0 = \Theta_0 = \Theta_1 = \Sigma_0 = 1$, then, the unconditional cost is given by

$$E\{u_0^2\} = (\frac{8}{33})^2 \cdot \frac{1}{2} \tag{B.12}$$

$$E\{v_0^2\} = (\frac{5}{33})^2 \cdot \frac{1}{2}) \tag{B.13}$$

$$E\{x_1^2\} = 2(1 - \frac{53\cdot 13}{(66)^2} \tag{B.14}$$

$$J_1 = 2 - \frac{187}{(22)(66)} \tag{B.15}$$

$$J_2 = \frac{3}{2} - \frac{589}{2(66)^2} \tag{B.16}$$

APPENDIX C

Consider the problem of obtaining a closed-loop equilibrium solution for the model in Section 3, using strategies with unconditional probability distribution dependence. The equivalent deterministic problem is

$$\Sigma_1 = \Sigma_0(1 + a + ab)^2 + \Xi_0(1+b)^2 a^2 + \Theta_0 \tag{C.1}$$

$$\Sigma_2 = \Sigma_1(1+c)^2 + \Theta_1 \tag{C.2}$$

$$J_1 = \Sigma_2 + c^2\Sigma_1 + a^2(\Sigma_0 + \Xi_0) \tag{C.3}$$

$$J_2 = \Sigma_2 + a^2b^2(\Sigma_0 + \Xi_0) \tag{C.4}$$

Using dynamic programming for the model of equations (C.1)-(C.4) yields

$$c^* = -\frac{1}{2} \tag{C.5}$$

$$b^* = -\frac{1}{5}\left(1 + \frac{\Sigma_0}{a(\Sigma_0+\Xi_0)}\right) \tag{C.6}$$

$$a^* = -\frac{8}{33}\,\frac{\Sigma_0}{\Sigma_0+\Xi_0} \tag{C.7}$$

This solution is exactly the same solution obtained in Appendix B.

Now consider the problem of finding an open-loop equilibrium. Using the equivalent deterministic model of equations (C.1) to (C.4), one obtains necessary conditions

$$c^* = \frac{1}{2} \tag{C.8}$$

$$\frac{\partial J_1}{\partial a} = 0 \tag{C.9}$$

$$\frac{\partial J_2}{\partial b} = 0 \tag{C.10}$$

$$\frac{\partial J_1}{\partial a} = 2(\Sigma_0+\Xi_0)a + (\Sigma_0(1+b)(1+a+ab) + \Xi_0(1+a)(a+ab) \tag{C.11}$$

$$\frac{\partial J_2}{\partial b} = a\,\frac{\Sigma_0}{2}\,(1+a+ab) + \frac{\Xi_0}{2}\,a(a+ab) + 2a^2b(\Sigma_0+\Xi_0) \tag{C.12}$$

Let $g = \frac{\Sigma_0}{\Sigma_0+\Xi_0}$. Then, (C.8)-(C.12) imply

$$2a + (1+b)^2 a + g(1+b) = 0 \tag{C.13}$$

$$4a^2b + a^2(1+b) + ga = 0 \tag{C.14}$$

Assuming $a \neq 0$, then

$$b = -\frac{1}{2} \pm\frac{\sqrt{3}}{2} \tag{C.15}$$

$$a = -\frac{4g}{33}\left(\frac{3}{2} \pm\sqrt{3}\right) \tag{C.16}$$

The solution costs, assuming $\Theta_0 = \Theta_1 = \Sigma_0 = \Xi_0 = 1$, are

$$J_1 = 2 - (\frac{1}{11})^2(\frac{27}{4} \pm 6\sqrt{3}) \quad (C.17)$$

$$J_2 = \frac{3}{2} - \frac{1}{2}(\frac{1}{11})^2(\frac{49}{12} \pm\frac{10\sqrt{3}}{3}) \quad (C.18)$$

REFERENCES

1. D. Castanon and N.R.Sandell,Jr.,"Equilibrium Solutions in Sequential Stochastic Games", IEEE Trans. on Auto. Control, Vol. AC-22, No. 4, August, 1977.
2. Y.C.Ho,"on the Minimax Principle and Zero-sum Stochastic Differential Games," JOTA, Vol. 13, No. 3, March 1974.
3. M. Simaan and J.B.Cruz Jr., "Additional Aspects of the Stackelberg Strategy in Nonzero-Sum Games," JOTA, Vol. 11, no.6,June 1973.
4. D. Castañon, "Equilibria in Stochastic Dynamic Games of Stackelberg Type," Report ESL-R-662, Electronics Systems Laboratory,MIT,May 1976.
5. S. Glankwamdee and J.B.Cruz, Jr., "Decentralized Stackelberg Strategies for Interconnected Stochastic Dynamic Systems," Report DC-1, Decision and Control Laboratory, CSL and EE,University of Illinois,April 1977.
6. J. B. Cruz, Jr., "Stackelberg Strategies for Multilevel Systems," in Directions in Large Scale Systems, Y. C. Ho and S. K. Mitter, editors, Plenum Publishing Corp., New York, 1976.
7. H. S. Witsenhausen, "A Standard Form for Sequential Stochastic Control," Mathematical Systems Theory, Vol. 7, 1973.
8. N. Sandell Jr., "On Open Loop and Closed Loop Nash Strategies," IEEE Trans. on Auto. Control, Vol. AC-19, June 1974.
9. M. Simaan and J.B.Cruz,Jr., "On the Stackelberg Strategy in Nonzero-Sum Games," JOTA, vol. 11, no. 5, 1973.