This final report summarizes the findings of research on the topic "Stochastic Dynamic Systems with Multiple Decision Makers and Parametric Uncertainties," supported by a Grant from the Air Force Office of Scientific Research, during the period May 1, 1988-August 31, 1989. The focus of the research during this 15-month period has been on the development of analytical and numerical solution techniques for stochastic control and team problems which involve active learning, the study of the impact of nonclassical information patterns in distributed decision making, and the development of real-time implementable distributed algorithms incorporating memory, for computation of equilibria in deterministic and stochastic games.
STOCHASTIC SYSTEMS WITH MULTIPLE DECISION MAKERS AND PARAMETRIC UNCERTAINTIES

AFOSR-88-0178

FINAL REPORT TO AFOSR
covering the period
May 1, 1988-August 31, 1989

Principal Investigator
Tamer Bašar

Decision and Control Laboratory
Coordinated Science Laboratory
University of Illinois
1101 W. Springfield Avenue
Urbana, Illinois 61801

December 28, 1989
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td></td>
</tr>
<tr>
<td>1. Multiple Decision-Maker Problems with Unknown Parameters</td>
<td>4</td>
</tr>
<tr>
<td>2. A Summary of Research Accomplishments</td>
<td>8</td>
</tr>
<tr>
<td>3. Publications Supported by the Grant (May 1988 - August 1989)</td>
<td>14</td>
</tr>
<tr>
<td>4. Conference Presentations and Plenary Talks</td>
<td>16</td>
</tr>
<tr>
<td>5. Attached Reprints and Preprints</td>
<td>17</td>
</tr>
</tbody>
</table>
ABSTRACT

This final report summarizes the findings of research on the topic "Stochastic Dynamic Systems with Multiple Decision Makers and Parametric Uncertainties", supported by a Grant from the Air Force Office of Scientific Research, during the period May 1, 1988 - August 31, 1989. The focus of the research during this 15-month period has been on the development of analytical and numerical solution techniques for stochastic control and team problems which involve active learning, the study of the impact of nonclassical information patterns in distributed decision making, and the development of real-time implementable distributed algorithms incorporating memory, for computation of equilibria in deterministic and stochastic games.
1. Multiple Decision-Maker Problems with Unknown Parameters

1.1. Solution concepts and some general approaches

The problem of strategic decision making in complex systems which involve multiple decision
makers (DM's), multiple objectives, and incomplete information arises frequently in the military con-
text. As compared with single DM problems, the analysis of multiple DM problems requires different
approaches and techniques, and furthermore certain standard features and properties we usually ascribe
to single DM problems do not generally extend naturally to multiple decision making. For example,
while, in single DM problems, optimization (minimization or maximization) of a single objective func-
tional would, in general, lead to a satisfactory decision policy (the so-called optimal policy), when the
decision problem involves multiple DM's and multiple objectives a plethora of possibilities emerge as
to the criterion which leads to a "satisfying" set of policies. Depending on the number of DM's, their
underlying goals, and the presence or absence of dominance in the decision making process, we may
have team-optimal, person-by-person optimal, Pareto optimal, Nash equilibrium, Stackelberg (leader-
follower) equilibrium, consistent conjectural variations equilibrium concepts, and several variants of
combinations of these in case of more than two DM's. Each of these leads, in general, to a different
outcome which is also a variant of the information structure of the problem (i.e., what each DM knows
a priori, what information he acquires during the evolution of the decision process, what information
exchange links are allowable, and what information transmission capability each DM is vested with).
The significance of information structure in multiple DM problems also manifests itself in the deriv-
ation of multimodel strategies: Model simplification through singular perturbations or aggregation is not
a well-posed procedure unless there is some kind of a matching between the information structures of
the original problem and the simplified version; no such inconsistencies arise, however, in single
decision-maker problems.
Until recently, most of the work on multiple DM problems has pertained to either deterministic systems or to systems with uncertain elements which have a complete probabilistic description—this a priori information being known by all the DM’s (the latter class of problems are also known as stochastic dynamic games). Hence, even though some decentralization of dynamically acquired information has been allowed for in the general formulation of dynamic games, it has been a common assumption to endow every DM with the common (centralized) a priori information regarding the complete statistical description of the “primitive” random variables. Our thesis has been that such an underlying assumption is not always a realistic one, especially when the decision problem involves distributed tasks for the DM’s. A more realistic formulation, in most cases, would allow for discrepancies in the perceptions of the DM’s regarding the underlying stochastic model. These discrepancies could be accommodated in the model by having a number of parameters which are either not stochastic or are stochastic but their complete statistical description is not known by all the DM’s.

The presence of unknown (or uncertain) parameters could affect the general problem formulation in basically three different ways:

i) **Through the objective functions.** Here, the objective function of the i’th DM may not be known completely by the j’th DM (j≠i), with the uncertainty characterized by a number of parameters whose values are unknown to the j’th DM.

ii) **Through the system response.** The evolution of the decision process may depend on a number of parameters whose values are unknown to some or all DM’s. [This type of uncertainty is also applicable to stochastic team problems.]

iii) **Through the measurements made by the DM’s.** Here either the observation scheme or the statistics of some of the variables in the measurement process of a DM (or both) may not be known to some other DM, with the uncertainty again being parameterized. [As in ii) this type of uncertainty is also applicable to stochastic team problems.]
Multiple DM problems with the types of uncertainties as described above can be treated by adopting essentially one of the following three approaches:

a) **Robustness or Minimum Sensitivity Approach.** Here one assumes some nominal values for the unknown parameters, determines a corresponding nominal performance for the system, and designs decision policies which would lead to minimum performance degradation should the parameters vary around their nominal values. The resulting decision policies are called *minimum sensitivity strategies*, and they are robust in a certain neighborhood of the nominal values. For some recent advances in this area and for motivation of this approach we refer to publications [P1],[P2] and [P11], which report our recent work on this topic, carried out under AFOSR support.

b) **Learning Schemes.** In this approach no nominal values for the unknown parameters will be available, but some *a priori* statistics may be attached to these parameters by the DM's, which will be updated in a decentralized manner as new dynamic information is acquired. This is akin to some of the methodologies developed earlier for control problems with unknown parameters (such as identification, parameter estimation, and adaptive control-still active research areas), which are, however, not applicable to multiple DM problems, because the rather intricate interactions of multiple DM's render any central learning scheme infeasible. The iterative schemes which are needed for such systems have to be decentralized and distributed, and have also to account for the possibility that DM's may not update their policies or actions in a predetermined order. There are the further questions of robustness of these schemes to possible inaccuracies in the computation phase during each update, and robustness to environmental changes. Learning schemes could involve two types of iterations: "iteration in the policy space" and "iteration in the decision (action) space". During the past year we have published papers on both types of iterations, as described in the next section.
7

c) **Minimax Approach.** Here no nominal values are available for the unknown parameters, but they are known to belong to some pre-specified sets. Then, the objective is to design strategies which would carry optimality or equilibrium property under worst possible values of the parameters on these sets. [See, for example, [P9] and [P10] for two different contexts where such a formulation would arise.] Such an approach entails a pessimistic design philosophy, and is applicable mostly to decision problems with a common objective functional (i.e., team problems). In multi-objective problems, the minimax philosophy may lead to some ambiguity, since what may seem to be a worst-case design for one objective functional may seem to lose this property when tested against a different objective functional. However, if different objective functionals are affected by different sets of unknown parameters, this approach would still be applicable. Furthermore, in some information transmission problems where the channel description is not complete, the minimax design (transmission) philosophy finds a natural home, as elucidated particularly in the recent paper [P5].

We should point out that a combination of any two or all three of the above approaches would also constitute a viable approach to multi-person decision problems with unknown parameters, which should be studied in proper contexts once the rudiments of a theory for each one separately is laid down.

1.2. **Stochastic teams: Nonclassical information patterns and computational algorithms**

A special class of problems which fall in the general framework outlined above is "stochastic teams," where there is necessarily a single objective, but the DM's (or agents) still use decentralized noise-corrupted measurements in the construction of their policies. If all the information is centralized, then the problem is no different from a stochastic control problem, which is relatively easier to analyze.
One of the most challenging classes of problems in multi-person decision making is the class of stochastic dynamic teams with *nonclassical information*. An information pattern is *nonclassical* if a decision maker’s (say, j’s) action affects the information of another DM (say, i), and there is no way in which DMi can infer the information based on which DMj’s action was determined. Under the nonclassical information pattern, the derivation of optimal team solution meets with formidable difficulties. One way of viewing these difficulties is that the control (or decision) plays a triple role, viz. (i) the deterministic control effort of reducing the error; (ii) improvement of future knowledge of uncertainty; (iii) signaling to agents acting in the future some useful information that they do not necessarily acquire; and these roles are in general conflicting. In spite of these difficulties, and the fact that such problems are (computationally) NP-complete, it may still be possible to identify classes of tractable problems with nonclassical information. Our research results indicate that this is indeed the case, as described in the next section.

2. A Summary of Research Accomplishments

In our proposal for this research, carried out under AFOSR support since May 1988, we had identified a number of challenging issues in multiple person decision making and control under uncertainty, dynamic games, and stochastic teams, and we outlined some research directions which need to be followed to enhance our understanding of the intricate role played by information patterns (especially patterns of nonclassical nature), active and passive learning, and hierarchies in such problems. We proposed to focus on both the methodological and algorithmic developments for the analysis and optimization of stochastic systems, and placed emphasis on the development of decentralized and distributed schemes using relaxation-type algorithms which incorporate memory.
Since May 1988, we have addressed several challenging issues in this context, and have made important strides. We provide below a brief summary of our research findings, full details of which can be found in the references listed in Section 3. Copies of all the references are attached (in full) to this report.

We now return to brief descriptions of the main contributions of the papers listed in Section 3.

Our first set of results pertain to stochastic teams with nonclassical information, and they contribute toward developing a general theory for multiple stage (finite and infinite horizon) stochastic optimization problems where the control action does not affect only the state trajectory but also the quality of information available to the decision makers, thus exhibiting a dual role. Two papers which report important results in that direction are [P2] and [P12], which deal with two totally different classes of stochastic decision problems with nonclassical information, and obtain explicit solutions in both cases using two totally different approaches.

In [P12] we consider a stochastic dynamic decision problem where at each step two consecutive decisions must be taken, one being what information bearing signal to transmit, and the other what control action to exert. Such a problem arises in the simultaneous optimization of both the observation and the control sequences in stochastic systems, and is a prime example of a stochastic team problem with nonclassical information. Using bounds from information theory, we were able to solve this problem completely for first-order systems under a quadratic cost criterion. We have shown in [P12] that, in the case of hard power constraints, the optimal measurement policy consists of transmitting the "innovation" in the new data at the maximum power level. In the case when the power levels at the transmitter are not fixed, the optimal power levels for transmitting this innovation can be found by solving a nonlinear optimal control problem. We also consider the case when the time horizon is infinite and the cost functional is discounted, in which context we prove the existence of optimal stationary control and transmission strategies, and provide a complete characterization of the optimal solu-
tion. Some versions of these results have also been presented in [P1] and [P3]. Furthermore, an extension has been provided in [P3], where we study stochastic systems of higher order, and show that the optimal design is generally nonlinear. We also provide a characterization of the best linear solution as well as an algorithm to compute it.

The problem treated in [P2] is another stochastic dynamic optimization problem which exhibits active learning, but the derivation of its solution requires a completely different approach. The proof of optimality given in the paper relates the original single objective problem to a sequence of nested zero-sum stochastic games. Existence of saddle points for these games implies the existence of optimal policies for the original stochastic control problem, which, in turn, can be obtained from the solution of a nonlinear deterministic optimal control problem. The paper also studies the problem of existence of stationary optimal policies when the time horizon is infinite and the objective function is discounted. This is one of the first reports in the literature on the derivation of closed-form solutions for a class of (non-neutral) stochastic control problems where the control directly affects the quality of information carried to future stages. In a subsequent work reported in [P4] we have provided a different perspective to the results of [P2] and [P12] by introducing a unifying framework and also analyzing the contrasts between the formulations and the method of solutions in the two papers.

In [P5], we consider the problem of optimum information transmission along channels with incomplete statistical description, and we seek a solution under a worst case analysis. We again operate under nonclassical information patterns and consider a number of cases depending on whether there are "hard" energy constraints or "soft" constraints on some decision variables and/or "soft" costs on communications. We obtain minimax decision rules in all these cases, some being saddle points and others not, the techniques of derivation being very much case-dependent. These are important prototype problems which should be considered essential building blocks for a general theory of multi-stage distributed decision making under nonclassical information, and with partial statistical description.
One paper in the list, which deals with the minimax philosophy is [P14], which addresses a fundamental problem in controller design, that of H-infinity optimization. Here we show, for the first time, that in the problem of disturbance attenuation in linear systems, there exist optimal nonlinear controllers which lead to a uniformly better performance (than that achievable by linear controllers) in a neighborhood of the worst-case disturbance. We also argue that a dynamic games set-up provides a natural framework for the time-domain derivation of optimal (minimax) controllers for tracking, model matching and disturbance rejection problems, in both finite and infinite horizons.

In paper [P6], we study the Nash equilibrium solution of continuous-time stochastic differential games with linear dynamics, when the decision makers (players) are allowed to develop different prior probabilities on the random variables appearing in the system dynamics and the measurements available to each player. We obtain precise conditions for the existence and uniqueness of noncooperative (Nash) equilibrium, and develop a method for iterative distributed computation of the corresponding policies. The distributed algorithm involves learning in the policy space, and it does not require that each player know the other's perceptions of the probabilistic model underlying the decision process. For the finite horizon problem, such an iteration converges whenever the length of the time horizon is short, and the limit in this case is an affine policy for all players if the underlying distributions are Gaussian. When the horizon is infinite, and a discount factor is used in the cost functionals, the iteration converges under conditions depending on the magnitude of the discount factor, the limiting policies again being affine in the case of Gaussian distributions.

In [P13] we study algorithms of the relaxation type, for the on-line, distributed computation of equilibria in noncooperative deterministic and stochastic games. These algorithms require memory on either the past policy choices or the past actions of the players, depending on whether the iteration (and learning) is in the policy or the action space. Our earlier work on this topic had shown that incorporating memory into Jacobi and Gauss-Seidel type iterations in two-person Nash games could lead to con-
siderable improvements in both the region of convergence and the speed of convergence towards non-cooperative equilibrium, under both asynchronous modes of update. In [P13] we show that these appealing features carry over to partially asynchronous modes of update, when there is a bound on the memory available to one or more players. We also study the robustness of the partially asynchronous relaxation-type algorithms to inaccuracies in each player's knowledge regarding the others' response functions.

In the two papers [P8] and [P9] we introduce and fully develop a new approach for policy optimization problems that involve the so-called "forward-looking" stochastic models. Such models provide a characterization of decision processes where the evolution of the underlying dynamics depends explicitly on the expectations the controlling agents form on the future evolution itself. They lead to non-standard stochastic dynamic optimization problems where one has to take into account the fact that there is a circular (closed) relationship between future forecasts and the future system behavior. In [P8] we study the class of models where the only input involves a two-step ahead prediction of the future system behavior, by formulating them as stochastic control problems (of the delayed information type) in both finite and infinite horizons. It is shown that when there is perfect state information, the solution is unique for both the finite and infinite horizon formulations, and it requires memory for the former while requiring only current state information for the latter. When only noisy measurements are available, it is shown that a certainty-equivalence type result holds, the memory requirements being the same as above, with perfect state now replaced by the output of a finite-dimensional filter. The second paper, [P9], extends these results to forward-looking models which have two types of controlled inputs -- a forecasting strategy and a tracking strategy. The objective of the second control input is to make the system track a given trajectory. This leads to a game-theoretic formulation, which we thoroughly study in [P9] for both finite and infinite horizons, and under both perfect and noisy state measurements. In all cases we show that the problem admits a unique Nash equilibrium solution, and provide a complete characterization of the corresponding decision rules.
In [P11] we consider a discrete zero-sum dynamic game which features delayed information and hence falls in the challenging class of games with partial information. Being a finite game, it admits a saddle-point solution in mixed strategies, but the brute-force approach toward construction of the saddle-point strategies fails since the complexity grows exponentially with the number of stages. Also, the game does not admit a time decomposition, as in feedback games, because of the delay in the acquisition of information by one of the players, which leads to "strong" correlation across stages. In the paper we develop a novel approach which "overlappingly decomposes" the original decision problem and enables the explicit computation of the saddle-point policies as well as the value of the game. The paper also discusses several extensions to scenarios which include noisy position measurements and/or extended memory for one of the players.

Finally, in [P10] we address the important issues of "time consistency" and "robustness" of equilibrium policies in noncooperative dynamic games, and introduce a general mathematical framework in which these issues can be given a precise meaning and analyzed effectively. The framework we develop enables us to obtain qualitative results independently of the underlying information pattern and the equilibrium solution adopted for the decision problem.

In a book to appear next year [B1], we have included a number of results on stochastic teams and games, which were generated as a part of the research efforts described above. It is unfortunate that the Grant was discontinued after one year of funding, leaving a number of important issues yet unresolved.
3. Publications Supported by the Grant (May 1988 - August 1989)

Books


Papers


Ph.D. Thesis

(Only those on topics related to the AFOSR Grant)


[C5] Paper [P10] was presented as a plenary talk at *Conference on Dynamic Games and Policy Making*, held at Tilburg University, the Netherlands, May 1988.


5. Attached Reprints and Preprints