Universität Bielefeld/IMW

Working Papers
Institute of Mathematical Economics

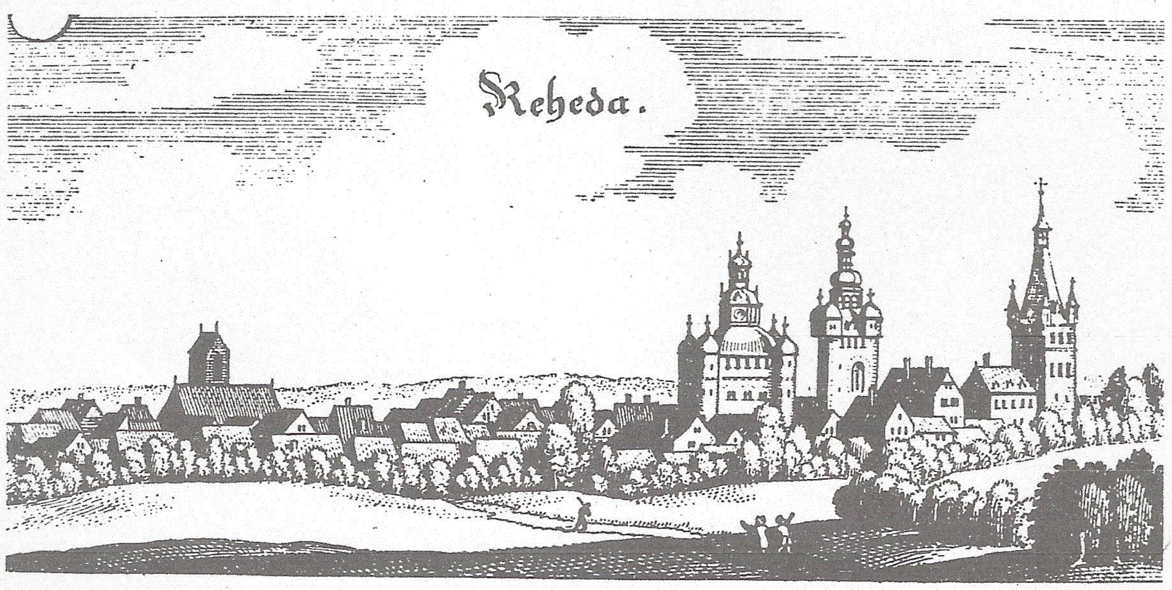
Arbeiten aus dem
Institut für Mathematische Wirtschaftsforschung

Nr. 26a

Hans W. Gottinger

Decomposition for a Static Stochastic Hierarchical System.

February 1975



Anno 1600 nach Merian

Institut für Mathematische Wirtschaftsforschung
an der
Universität Bielefeld
Adresse/Address:
Schloß Rheda
484 Rheda
Bundesrepublik Deutschland
Federal Republic of Germany

DECOMPOSITION FOR A STATIC STOCHASTIC HIERARCHICAL SYSTEM.

Hans W. Gottinger

Abstract:

The decentralized control of a static stochastic large-scale system is considered. Particular emphasis is given to control strategies which utilize decentralized information and can be computed in a decentralized manner.
The deterministic constrained optimization problem is generalized to the stochastic case when each decision variable depends on different information and the constraint is only required to be satisfied on the average. For problems with a particular structure, a hierarchical decomposition is obtained.

DECOMPOSITION FOR A STATIC STOCHASTIC HIERARCHICAL SYSTEM.

1. Introduction

In this paper we consider the stochastic optimization problem of a static system consisting of several subsystems. Each subsystem has a decision agent which has noisy information on the state of the system. The overall objective of the system is the sum of individual objectives of the subsystems. The subsystems are uncoupled except for constraints, which couple them together. Contrary to the deterministic case, the constraints do not have to be satisfied exactly. Rather, the problem solver only requires the constraints to be satisfied on the average. We have thus a constrained stochastic optimization problem with several decision agents each having noisy and different information on the state. The many decision agent aspect of the problem has been considered under the heading of team theory [8]. For a constrained deterministic problem with the special structure described above, a hierarchical decomposition has been obtained using mathematical programming [5] [6]. We shall consider the two aspects of the problem simultaneously and obtain a hierarchical decomposition. This static problem is not only interesting for its own sake but is also useful for the decomposition of dynamic systems.

In the next section we present an example to motivate the general problem that we will study in this paper. In Section 3 we review some results in non-linear programming; these can be used to obtain the decomposition of a static optimization problem when the state of the system is observed exactly. In Section 4 the stochastic optimization problem is formulated for the case when the state of the system is not known exactly. In Section 5 the decomposition of the stochastic problem is investigated. Conditions under which the decomposition is well-posed are given and related to the information structure of the system. In Section 6 these results are stated in terms of measurement functions. The stochastic version of the example is solved in Section 7 and compared with the deterministic solution.

2. An Example

Consider a manufacturing company with N divisions, each producing a set of different commodities using the same resources. The ith division produces $\underline{u}_i$ units of goods $\underline{G}_i$ from $\underline{A}_i\underline{u}_i$ units of raw material at a cost of $\underline{u}_i'\underline{R}_i\underline{u}_i$ where $\underline{R}_i$ is assumed to be a positive definite matrix.

The market price of $\underline{G}_i$ is $2\underline{\pi}_i$ and the total resources available are $\underline{v}$.

Given any price vector $2\underline{\pi}_i$ and production $\underline{u}_i$, the profit function of the ith division is

$$- f_i(\underline{u}_i,\underline{\pi}_i) = 2\underline{u}_i'\underline{\pi}_i - \underline{u}_i'\underline{R}_i\underline{u}_i \qquad (2.1)$$

The total profit of the company is the sum of the profits of all the divisions, i.e.,

$$-f(\underline{u},\underline{\pi}) = - \sum_{i=1}^{N} f_i(\underline{u}_i,\underline{\pi}_i) \qquad (2.2)$$

The objective of the company is to minimize the total loss (maximize the total profit) subject to the constraint that the total resources used are less than the total resources available. The problem is thus

Problem 1: Minimize over $\underline{u}_1,\ldots,\underline{u}_N$

$$\sum_{i=1}^{N} \underline{u}_i'\underline{R}_i\underline{u}_i' - 2\underline{u}_i'\underline{\pi}_i \qquad (2.3)$$

$$\sum_{i=1}^{N} \underline{A}_i\underline{u}_i - \underline{v} \leq \underline{0} \qquad (2.4)$$

Remark: We could have imposed the additional constraint that $\underline{u}_i \geq \underline{0}$, i=1,....., N but for simplicity we have assumed implicitly that the $\underline{u}_i$'s would turn out to be non-negative when Problem 1 is solved.

In this example the state of the system consists of the price vector $\underline{\pi}_i$, i=1,...., N, the resource vector $\underline{v}$ and possibly the cost matrices $\underline{R}_i$ and the resource utilization matrices $\underline{A}_i$. The decisions to be chosen are $\underline{u}_i$, i=1,...., N. Calling the state as x we have the following general problem

Problem 2:

$$\text{Minimize} \sum_{i=1}^{N} f_i(u_i,x)$$

$$\text{Subject to} \sum_{i=1}^{N} g_i(u_i,x) - g_o(x) \underline{\leq} \underline{0} \qquad (2.5)$$

For our example

$$f_i(u_i,x) = \underline{u}_i'\underline{R}_i\underline{u}_i - 2\underline{u}_i'\underline{\pi}_i \qquad (2.6)$$

$$g_i(u_i,x) = \underline{A}_i\,\underline{u}_i \qquad (2.7)$$

$$g_o(x) = \underline{v} \qquad (2.8)$$

There are situations when the state of the system cannot be observed exactly, but is described probabilistically. Suppose now that $\underline{\pi}_i$ is measured by the ith division manager as

$$\underline{z}_i = \underline{C}_i\,\underline{\pi}_i + \underline{\theta}_i \qquad i=1,....,N \qquad (2.9)$$

$\underline{v}$ is measured by the resource manager as

$$\underline{z}_0 = \underline{C}_0\,\underline{v} + \underline{\theta}_0 \qquad (2.10)$$

$\underline{\pi}_i$, $\underline{\theta}_i$, i=1,.... N, $\underline{v}$ and $\underline{\theta}_0$ are random vectors independent of each other and having the normal distributions (assumed known)

$$E\{\underline{\pi}_i\} = \overline{\underline{\pi}}_i \quad ; \quad Var\{\underline{\pi}_i\} = \underline{\Pi}_i \qquad (2.11)$$

$$E\{\underline{\theta}_i\} = \underline{0} \quad ; \quad Var\{\underline{\theta}_i\} = \underline{\Theta}_i \qquad i=1,\ldots, N \tag{2.12}$$

$$E\{\underline{v}\} = \overline{\underline{v}} \quad ; \quad Var\{\underline{v}\} = \underline{V} \tag{2.13}$$

$$E\{\underline{\theta}_0\} = \underline{0} \quad ; \quad Var\{\underline{\theta}_0\} = \underline{\Theta}_0 \tag{2.14}$$

All the information available are contained in the measurements $\underline{z}_i$, $i=0,\ldots, N$. The production of each division has to be based on his measurement and some other signal based on $\underline{z}_0$.

The objective of the company is to minimize the expected total loss. As for the resource constraint (2.4) it can no longer be satisfied exactly since $\underline{v}$ is not measured exactly. Instead, we require the total resources used to be less than the total resources available given the measurement $\underline{z}_0$, i.e.

$$E\left\{\sum_{i=1}^{N} \underline{A}_i \underline{u}_i - \underline{v} \mid \underline{z}_0\right\} \leq \underline{0} \tag{2.15}$$

The production of each division has to use some information contained in $\underline{z}_0$ because the resource constraint (2.15) has to be satisfied. We thus have the following problem,

<u>Problem 1A:</u> Minimize $E\left\{\sum_{i=1}^{N} \underline{u}_i'\underline{R}_i\underline{u}_i - 2\underline{u}_i'\underline{\pi}_i\right\}$ (2.16)

subject to

$$E\left\{\sum_{i=1}^{N} \underline{A}_i \underline{u}_i - \underline{v} \mid \underline{z}_0\right\} \leq \underline{0} \tag{2.17}$$

$$\underline{u}_i = \underline{\eta}_i(\underline{z}_i ; \underline{z}_0) \qquad i=1,\ldots, N \tag{2.18}$$

<u>Remark:</u> $\underline{u}_i$ at most can depend on all the information contained in $\underline{z}_i, \underline{z}_0$. We shall show later that the optimal decision function in some cases can

be found in a hierarchical manner and operation of the company can be decentralized.

3. Decomposition of a Non-Linear Programming Problem

In this section, we present some results in non-linear programming. These give rise immediately to a decomposition method for deterministic problems. Later on they will be used to obtain a decomposition for the stochastic case.

Consider the mathematical programming problem.

Problem 3: Minimize $f(u_1, \ldots, u_N)$

$$\text{Subject to } g(u_1, \ldots, u_N) \leq \underline{0} \in R^p \tag{3.1}$$

$$u_i \in U_i \qquad i=1, \ldots, N$$

where

$$f(u_1, \ldots, u_N) = f_1(u_1) + \ldots\ldots + f_N(u_N) \tag{3.2}$$

$$g(u_1, \ldots, u_N) = g_1(u_1) + \ldots\ldots + g_N(u_N) - g_o \tag{3.3}$$

Except for the coupling constraint (3.1), the problem is essentially uncoupled. The constraint may be interpreted as the common resource available to all the decision makers. This structure has been exploited to give a hierarchical decomposition scheme for the solution of the problem using results in mathematical programming. We state one sufficient condition which makes this possible.

Theorem 3.1 (Saddle-point condition): Let f be a real-valued function defined on a subset C of a linear space U. Let g be a mapping from C into the Euclidean space R^p. Suppose there exists a $\underline{p}^* \in R^p$, $\underline{p}^* \geq \underline{0}$ and a $u^* \in C$ such that the Lagrangian $L(u,\underline{p}) \triangleq f(u) + \underline{p}'g(u)$ possesses a saddle-point at $u^*,\underline{p}^*$, i.e.,

$$L(u^*,\underline{p}) \leq L(u^*,\underline{p}^*) \leq L(u^*,\underline{p}^*) \qquad (3.4)$$

for all $u \in C$, $\underline{p} \geq \underline{0}$ then u^* solves

$$\text{minimize } f(u)$$

$$\text{Subject to } g(u) \leq \underline{0} \qquad u \in C \qquad (3.5)$$

The proof of this theorem is elementary [7]. Note that there are no conditions on the convexity or differentiability of f or g. For equality constraints, the same result holds except that $\underline{p}$ is no longer required to be non-negative. The following theorem is due to Lasdon [5].

Theorem 3.2: Suppose there exists a saddlepoint for the Lagrangian corresponding to Problem 3, then the following hierarchical scheme can be used to obtain a solution, provided the minimizing problem is well-posed.*

Lower level: $$\text{Minimize } \tilde{L}_i(u_i,\underline{p}) = f_i(u_i) + \underline{p}' g_i(u_i)$$

$$\text{Subject to } u_i \in U_i$$

$$i=1,\ldots, N \qquad (3.6)$$

Higher level: $$\text{Maximize } \sum_{i=1}^{N} \tilde{L}_i^*(\underline{p}) - \underline{p}' g_0$$

$$\text{Subject to } \quad \underline{p} \geq \underline{0} \qquad (3.7)$$

where $\tilde{L}_i^*(\underline{p})$ is the minimum obtained in equation (3.6).

---

*For some $\underline{p}$, the lower level problem may not have a solution. We thus have to limit $\underline{p}$ to the set $D = \{\underline{p} \mid \text{the lower level problem has a solution}\}$.

Proof: We need the fact that the constrained saddle-point for L(a,b), $a \varepsilon A$, $b \varepsilon B$ exists if and only if [10]

$$\underset{a \varepsilon A}{\text{Min}} \; \underset{b \varepsilon B}{\text{Max}} \; L(a,b) = \underset{b \varepsilon B}{\text{Max}} \; \underset{a \varepsilon A}{\text{Min}} \; L(a,b) \qquad (3.8)$$

The value of the saddle-point is also equal to either side of equation (3.8). Given any $\underline{p}$ we note that the minimization part on the right side of equation (3.8) can be split up into N minimization problems independent of each other. Specifically, we have

$$\underset{\underline{p} \geq \underline{0}}{\text{Max}} \; \underset{u}{\text{Min}} \; L(u,\underline{p}) = \underset{\underline{p} \geq \underline{0}}{\text{Max}} \; \underset{u}{\text{Min}} \left\{ \sum_{i=1}^{N} f_i(u_i) + \sum_{i=1}^{N} \underline{p}' g_i(u_i) - \underline{p}' g_o \right\}$$

$$= \underset{\underline{p} \geq \underline{0}}{\text{Max}} \; \sum_{i=1}^{N} \underset{u_i}{\text{Min}} \{ f_i(u_i) + \underline{p}' g_i(u_i) \} - \underline{p}' g_0 \qquad (3.9)$$

Equations (3.6) and (3.7) are obtained by making the appropriate identifications. Q.E.D.

Theorem 3.2 suggests a way of finding the optimal $\underline{p}^*$ and u^* simultaneously. This requires giving $\tilde{L}_i^*(\underline{p})$ as a function of $\underline{p}$. There are numerical methods [5] by which the optimal solution is found recursively by choosing a new $\underline{p}_{t+1}$ depending on the result of optimizing the dual function $\sum_{i=1}^{N} \tilde{L}_i(u_i, \underline{p}_t)$. However, we are more interested in the structure of the decomposition, i.e., once an optimal $\underline{p}^*$ is found, the lower level problems are uncoupled. The significance of this is more obvious when we look at the parametric case given by Problem 2. For each x we have a mathematical programming problem; x may be regarded as the state of the system which is known exactly. If we use the result of Theorem 3.2, the optimal $\underline{p}^*$ would be a function of x, i.e., $\underline{p}^*(x)$. With this optimal $\underline{p}^*(x)$, the lower level problems would be

$$\text{Minimize } \tilde{L}_i(u_i,\underline{p}^*(x),x) = f_i(u_i,x) + \underline{p}^*(x)g_i(u_i,x)$$

$$u_i \in U_i \qquad i=1,\ldots,N \quad (3.11)$$

Thus we can regard the higher level and lower level decision makers as both making observations on the system. The higher level decision maker (coordinator) observes the state x, chooses the coordinating parameter $\underline{p}^*(x)$ and transmits it to the lower level. The lower level decision makers then use this, together with f_i and g_i and x to choose their optimal decisions. This is displayed in fig. 1.

Applying this result to the example given in Problem 1 we have the following decomposition:

Lower level (Division manager):

$$\underset{\underline{u}_i}{\text{Minimize}} \ \underline{u}_i'\underline{R}_i\underline{u}_i - 2\underline{u}_i'\underline{\pi}_i + \underline{p}'\underline{A}_i\underline{u}_i \qquad (3.12)$$

$$i=1,\ldots,N$$

Denote the optimal of (3.12) by $\tilde{L}_i^*(\underline{p})$

Higher level (Resource manager):

$$\underset{\underline{p}\geq\underline{0}}{\text{Max}} \ \sum_{i=1}^{N} \tilde{L}_i^*(\underline{p}) - \underline{p}'\underline{v} \qquad (3.13)$$

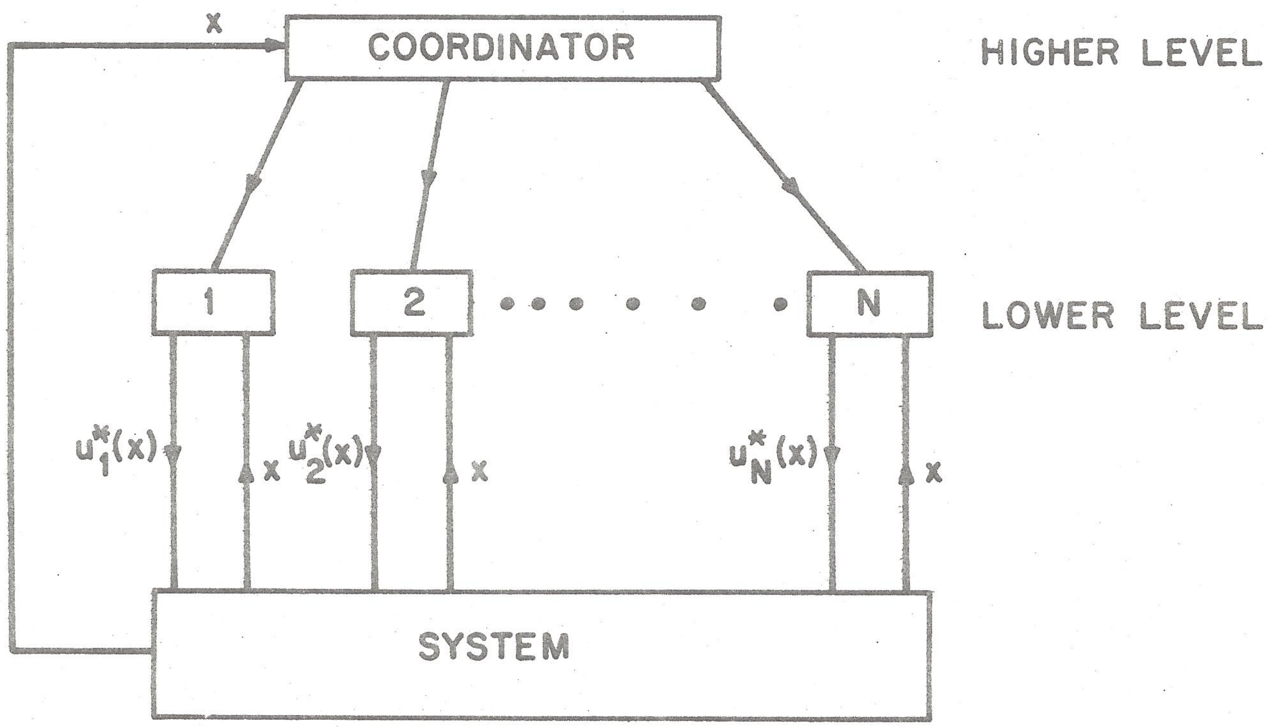


Fig. 1 Structure of Decomposition (Deterministic)

From these equations we obtain the following optimal $\underline{u}_i^*, i=1, \ldots N$ and $\underline{p}^*$

$$\underline{u}_i^* = \underline{R}_i^{-1} [\underline{\pi}_i - \tfrac{1}{2}\underline{A}_i'\underline{p}^*] \qquad (3.14)$$

$$\underline{p}^* = \underset{\underline{p} \geq \underline{0}}{\text{Arg Max}} - \tfrac{1}{4}\underline{p}'\Big(\sum_{i=1}^{N} \underline{A}_i\underline{R}_i^{-1}\underline{A}_i'\Big)\underline{p} + \underline{p}'\Big(\sum_{i=1}^{N} \underline{A}_i\underline{R}_i^{-1}\underline{\pi}_i - \underline{v}\Big) - \Big(\sum_{i=1}^{N} \underline{\pi}_i'\underline{R}_i^{-1}\underline{\pi}_i\Big) \qquad (3.15)$$

Referring to equation (3.12) we see that the loss function of the ith division manager has been modified by the addition of an extra term which reflects the cost of resources. $\underline{p}$ is the price of the resources while $\underline{A}_i\underline{u}_i$ denotes the amount used.

In this deterministic case, the lower level decision makers base their decisions on $\underline{\pi}_i$ while the higher level bases his decision on $\underline{\pi}_i$ and $\underline{v}$. There is some decentralization of information, but the higher level in fact needs more information than the lower level. In the general deterministic case, both levels need the same information x, which is not too satisfactory. This leads us to study the stochastic case when information can also be decentralized.

4. Formulation of the Stochastic Problem

We now consider the case when the state x is not known exactly by the different decision makers. However, there is a probability description on the state space X given by the triplet $(X,\mathcal{B},\mu)$. $\mathcal{B}$ is a σ- algebra on X, and μ is a probability measure.

Let $\mathcal{F}_i$, $i=1,\ldots,N$ be sub-σ-algebras of $\mathcal{B}$. $\mathcal{F}_i$ represents the information available to the ith decision maker. Since the state x is not observed exactly, u_i will be required to be generated by a function γ_i measurable with respect to $\mathcal{F}_i$. This is equivalent to the existence of a measurement function h_i on x such that u_i depends on the measurement $z_i = h_i(x)$ [4]. Denote by Γ_i the set of admissible decision functions γ_i measurable with respect to $\mathcal{F}_i$. Then $\gamma \triangleq (\gamma_1,\ldots,\gamma_N) \in \Gamma_1 \times\ldots\times \Gamma_N \triangleq \Gamma$. Given any decision function γ, $f(\gamma(x),x)$ would be a random variable. As in the case of team decision problems γ is chosen to minimize the expected payoff $E\{f(\gamma(x),x)\}$.

For the constraint several alternative formulations are possible.

1. $g(\gamma(x),x) \leq 0$ a.e. (4.1)

   As would be expected, it is rather difficult to satisfy this constraint.

2. $\text{Prob}\,\{g(\gamma(x),x) \leq 0\} \geq b$ (4.2)

   where b is some given probability.

Particular cases of this problem have been studied under the heading of <u>chance constrained programming</u> [1]. It is the situation where the constraint is only required to be satisfied with a given probability.

3. $E\{g(\gamma(x),x)|F_0\} \leq 0$ a.e. (4.3)

where F_0 is some sub-σ-field of B. F_0 specifies the degree of exactness with which the constraint has to be satisfied or in other words the information of a coordinator who sees that the constraint is satisfied.

Two extreme cases are possible:

a. $F_0 = \{\Phi, X\}$ (4.4)

This corresponds to no measurements for the coordinator. Then

$$E\{g(\gamma(x),x)\} \leq 0 \qquad (4.5)$$

b. $F_0 = B$ (4.6)

This corresponds to measuring the state almost exactly. Then

$$g(\gamma(x),x) \leq 0 \quad \text{a.e.} \qquad (4.7)$$

With the introduction of the constraint, the information available to the decision makers may not be sufficient to insure that the constraint is satisfied. In general some extra information has to be communicated from the coordinator to the decision makers.

We will investigate what this information should be. Let $\Gamma_i' \supset \Gamma_i$ be the new admissible functions. Γ_i' is set of functions measurable with respect to $F_i \cap F_0$. Thus we have formulated the following stochastic optimization problem.

Problem 4: Minimize $E\{f(\gamma(x),x)\}$

Subject to $E\{g(\gamma(x),x)|F_0\} \leq 0$ a.e.

$$\gamma = (\gamma_1,\ldots,\gamma_N) \in \Gamma_1' \times \ldots \times \Gamma_N'$$

$$f(\gamma(x),x) = f_1(\gamma_1(x),x) + \ldots + f_N(\gamma_N(x),x)$$

$$g(\gamma(x),x) = g_1(\gamma_1(x),x) + \ldots + g_N(\gamma_N(x),x) - g_o(x) \qquad (4.8)$$

Remark: Γ_i' is the set of decision functions which use both the information of the ith decision maker as well as the information of the coordinator. We shall show later that not all the information of the coordinator is needed by the ith decision maker to choose his best decision. Under certain conditions, the information of the coordinator can be compressed into a signal which will be sufficient for the ith decision maker.

5. Decomposition of the Stochastic Problem

The special form of the constraint allows us to transform Problem 4 into a simpler form for which the results of section 3 are applicable.

Lemma 5.1: Let $f(\gamma(x),x)$ be a random function from $\Gamma' \times X$ into the reals, where Γ' is a set of functions on X measurable with respect to $F \cap F_0$. $F \subset B$ and $F_0 \subset B$. Γ is the set of functions measurable with respect to F.

Let $$M = \{\gamma | E\{g(\gamma(x),x)|F_0\} \leq 0 \text{ a.e.}\}$$

Suppose $\min\limits_{\gamma(\cdot;y) \in \Gamma \cap M} E\{f(\gamma(x;y),x)|F_0\}(y)$ exists a.e. and is equal

to $E\{f(\gamma^*(x;y),x)|F_0\}(y)$, then

$$\min_{\gamma \in \Gamma' \cap M} E\{f(\gamma(x),x)\} = E\{f(\gamma^*(x;x),x)\}$$
$$= E\{ \min_{\gamma(\cdot;y) \in \Gamma \cap M} E\{f(\gamma(x;y),x)|F_0\}(y)\} \quad (5.1)$$

Proof: For $\gamma(\cdot) \in \Gamma' \cap M$ $\gamma(\cdot;y) \in \Gamma \cap M$

$$E\{f(\gamma(x;y),x)|F_0\}(y) = E\{f(\gamma(x),x)|F_0\}(y) \quad (5.2)$$

For a proof of this see Appendix A.

Thus

$$\min_{\gamma(\cdot;y) \in \Gamma \cap M} E\{f(\gamma(x;y),x)|F_0\}(y) = E\{f(\gamma^*(x;y),x)|F_0\}(y)$$
$$\leq E\{f(\gamma(x),x)|F_0\}(y) \text{ a.e. for all } \gamma \in \gamma' \cap M \quad (5.3)$$

Taking the unconditional expectation and minimizing over $\Gamma' \cap M$ we have,

$$E\{\min_{\gamma(\cdot;y)\,\varepsilon\,\Gamma\cap M} E\{f(\gamma(x;y),x)|F_0\}(y)\} \leq \min_{\gamma\,\varepsilon\,\Gamma'\cap M} E\{f(\gamma(x),x)\} \quad (5.4)$$

On the other hand

$$E\{\min_{\gamma(\cdot;y)\,\varepsilon\,\Gamma\cap M} E\{f(\gamma(x;y),x)|F_0\}(y)\} = E\{f(\gamma^*(x),x)\} \geq \min_{\gamma\,\varepsilon\,\Gamma'\cap M} E\{f(\gamma(x),x)\} \quad (5.5)$$

From equations (5.4) and (5.5) we obtain equation (5.1). Q.E.D.

Using Lemma 5.1, Problem 4 can be solved by considering the following problem.

Problem 5: Minimize $E\{f(\gamma(x;y),x)|F_0\}(y)$ a.e.

Subject to $E\{g(\gamma(x;y),x)|F_0\}(y) \leq 0$ a.e. (5.6)

$$\gamma(\cdot;y)\ \varepsilon\ \Gamma_1 x \ldots x\Gamma_N$$

If F_0 is such that the conditional probability measure $P_y{}^\circ(A)$ is regular, i.e. it is a probability measure given any y, then Problem 5 can be transformed to the following form.

Problem 6: Minimize $\hat{f}(\gamma;y)$

Subject to $\hat{g}(\gamma;y) \leq 0$

$$\gamma(\cdot;y)\ \varepsilon\ \Gamma \quad (5.7)$$

where $$\hat{f}(\gamma;y) \triangleq E\{f(\gamma(x;y),x)|F_0\}(y) = \int f(\gamma(x;y),x)\,dP_y{}^\circ(x) \quad (5.8)$$

$$\hat{g}(\gamma;y) \triangleq E\{g(\gamma(x;y),x)|F_0\}(y) = \int f(\gamma(x;y),x)\,dP_y{}^\circ(x) \quad (5.9)$$

Remark: The conditional probability measure is regular if it is generated by an observation function [2].

Problem 6 is a conventional functional minimization problem given any y. The results in Theorems 3.1 and 3.2 do not depend on the finite dimensionality of u, thus a hierarchical decomposition is obtained if a saddle-point exists for Problem 5. This is summarized in the following theorem.

<u>Theorem 5.2</u>: Suppose there exists a saddle-point $(\gamma^*(\cdot;y), p^*(y))$ for the Lagrangian associated with Problem 5. Then Problem 5 can be solved by the following hierarchical decomposition.

<u>Lower level</u>:

$$\text{Minimize } \tilde{L}_i(\gamma_i(\cdot;y), \underline{p}(y), y) = E\{f_i(\gamma_i(x;y),x) + \underline{p}'(y) g_i(\gamma_i(x;y),x) \mid F_0\}(y)$$

$$\gamma_i(\cdot;y) \in \Gamma_i \qquad i=1,\ldots\ldots, N \qquad (5.10)$$

<u>Higher level</u>:

$$\text{Maximize } \sum_{i=1}^{N} \tilde{L}_i^*(\underline{p}(y), y) - E\{\underline{p}'(y) g_0(x) \mid F_0\}(y)$$

$$\underline{p}(y) \geq 0 \qquad (5.11)$$

where $\tilde{L}_i^*(\underline{p}(y),y)$ is the minimum obtained in equation (5.10).

<u>Proof</u>: By using Theorem 3.2 on Problem 6, the decomposition is obtained.

Corresponding to Problem 4 we have the following decomposition.

<u>Higher level</u>: Choose $\underline{p}^*(y)$ measurable with respect to F_0.

<u>Lower level</u>:

$$\text{Minimize } \tilde{L}_i(\gamma_i(\cdot;y), \underline{p}^*(y), y) = E\{f_i(\gamma_i(x;y),x) + \underline{p}^{*\prime}(y) g_i(\gamma_i(x;y),x) \mid F_0\}(y)$$

$$\gamma_i(\cdot;y) \in \Gamma_i \qquad i=1.\ldots\ldots, N \qquad (5.12)$$

Note the optimal γ_i^* can be expressed in the form $\gamma_i^*(x, p^*(x))$.

The optimization problem of each lower level decision maker is described by equation (5.12). A conditional expectation has to be optimized by each. This optimization is not always well-defined with the information available to the ith decision maker. We give a necessary and sufficient condition when this is defined.

Theorem 5.3: Let $\mathcal{G}_i$ be the smallest σ - algebra of $\mathcal{F}_0$ with respect to which $E\{f_i(\gamma_i(x;y),x) + \underline{p}'(y)g_i(\gamma_i(x;y),x)|\mathcal{F}_0\}$ is measurable. Then given $\underline{p}(y)$, $L_i(\gamma_i(\cdot;y),\underline{p}(y),y)$ can be optimized by the ith decision maker if and only if $\mathcal{G}_i \subset \mathcal{F}_i$.

Proof: For any measurable function $\ell(x)$, if $E\{\ell|\mathcal{F}_0\}$ is measurable with respect to $\mathcal{G}_i$, $\mathcal{G}_i \subset \mathcal{F}_0$, then $E\{\ell|\mathcal{F}_0\} = E\{\ell|\mathcal{G}_i\}$. (see Appendix A) If $\mathcal{G}_i \subset \mathcal{F}_i$, then

$$E\{f_i(\gamma_i(x;y),x) + \underline{p}'(y)g_i(\gamma_i(x;y),x)|\mathcal{F}_0\}$$

$$= E\{f_i(\gamma_i(x;y),x) + \underline{p}'(y)g_i(\gamma_i(x;y),x)|\mathcal{G}_i\}$$

$$= E\{E\{f_i(\gamma_i(x;y),x) + \underline{p}'(y)g_i(\gamma_i(x;y),x)|\mathcal{F}_i\}|\mathcal{G}_i\} \qquad (5.13)$$

The inner expectation can be evaluated by the ith agent and minimized with respect to $\gamma_i(.;y) \in \Gamma_i$, hence minimizing $\tilde{L}_i(\gamma_i(\cdot;y),\underline{p}(y),y)$. If $\mathcal{G}_i \not\subset \mathcal{F}_i$, then $E\{f_i(\gamma_i(x;y),x) + \underline{p}'(y)g_i(\gamma_i(x;y),x)|\mathcal{G}_i\}$ cannot be evaluated given the information contained in $\mathcal{F}_i$, and thus it cannot be minimized. Q.E.D.

$\mathcal{G}_i$ represents the minimal sufficient information required by the ith agent to solve the decomposed decision problem given only $\underline{p}(y)$. If this information is not available, then the coordinator has to supply something else besides $\underline{p}(y)$. Typically this would be $P_y^i(A)$, the conditional probability measure with respect to $\mathcal{G}_i$. Note that although $\mathcal{F}_0 \subset \mathcal{F}_i$ satisfies the condition in Theorem 5.3, it is not always necessary for the ith agent to have more information than the coordinator. This will be illustrated in the next section.

6. Reformulation in Terms of Measurement Functions

In order to gain more insight, we shall reformulate the problem in terms of probability densities and measurement functions. The information requirements for the hierarchical decomposition can then be seen more easily.

Let x be the state of the system. x includes noises as well.

$z_i = h_i(x)$ be the measurement of the ith agent; $z_i \in Z_i$

$z_0 = h_0(x)$ be the measurement of the coordinator (specifying the constraint); $z_0 \in Z_0$

Then F_i, i=1,...., N is the σ - field on X generated by h_i and γ_i is measurable with respect to F_i if $\gamma_i = \eta_i o h_i$ where η_i is Borel-measurable on Z_i.

Corresponding to Problem 4 we have

Problem 7: Minimize $E\{f(\eta(z),x)\}$

Subject to $E\{g(\eta(z),x)|z_0\} \leq 0$

$$\eta(z) = (\eta_1(z_1;z_0),.....\eta_N(z_N;z_0))$$

$$f(\eta(z),x) = f_1(\eta_1(z_1;z_0),x) + ... + f_N(\eta_N(z_N;z_0),x)$$

$$g(\eta(z),x) = g_1(\eta_1(z_1;z_0),x) + ... + g_N(\eta_N(z_N;z_0),x) - g_0(x) \quad (6.1)$$

Corresponding to Problem 5, we have

Problem 8: Minimize $E\{f(\eta(z),x)|z_0\}$

Subject to $E\{g(\eta(z),x)|z_0\} \leq 0$

with η, f and g given as in equation (6.1) (6.2)

Theorem 5.2 then becomes

Theorem 6.1: Suppose there exists a saddle-point $(\eta^*(\cdot;z_0),\underline{p}^*(z_0))$ for the Lagrangian associated with Problem 8, then Problem 8 can be solved by the following hierarchical decomposition.

Lower level:

Minimize

$$\tilde{L}_i(\eta_i(\cdot;z_0),\underline{p}(z_0),z_0) = E\{f_i(\eta_i(z_i;z_0),x) + \underline{p}'(z_0)g_i(\eta_i(z_i;z_0),x)\,|\,z_0\}$$

$$i=1,\ldots\ldots, N \qquad (6.3)$$

Higher level:

$$\text{Maximize} \sum_{i=1}^{N} \tilde{L}_i^*(\underline{p}(z_0),z_0) - E\{\underline{p}'(z_0)g_0(x)\,|\,z_0\}$$

$$\text{Subject to } \underline{p}(z_0) \geq \underline{0} \qquad (6.4)$$

$\tilde{L}_i^*(\underline{p}(z_0),z_0)$ is the minimum obtained in equation (6.3).

Remark: From equation (6.3) we conclude that $\eta_i^*(z_i;z_0) = \eta_i^*(z_i;\underline{p}^*(z_0))$, i.e., all the relevant information about the constraint is contained in $\underline{p}^*(z_0)$ if the lower level problem is well defined.

The hierarchical decomposition scheme for Problem 7 then consists of the following.

Higher level: Coordinator makes a measurement z_0, computes the coordinating parameter $\underline{p}^*(z_0)$ and sends it to the lower level.

Lower level: ith decision agent makes a measurement z_i, and uses this together with $\underline{p}^*(z_0)$ to compute the best decision function $\eta_i^*(z_i;\underline{p}^*(z_0))$.

The structure of the decomposition is displayed in Figure 2. Note that the decomposition is in real-time since no iterations are involved.

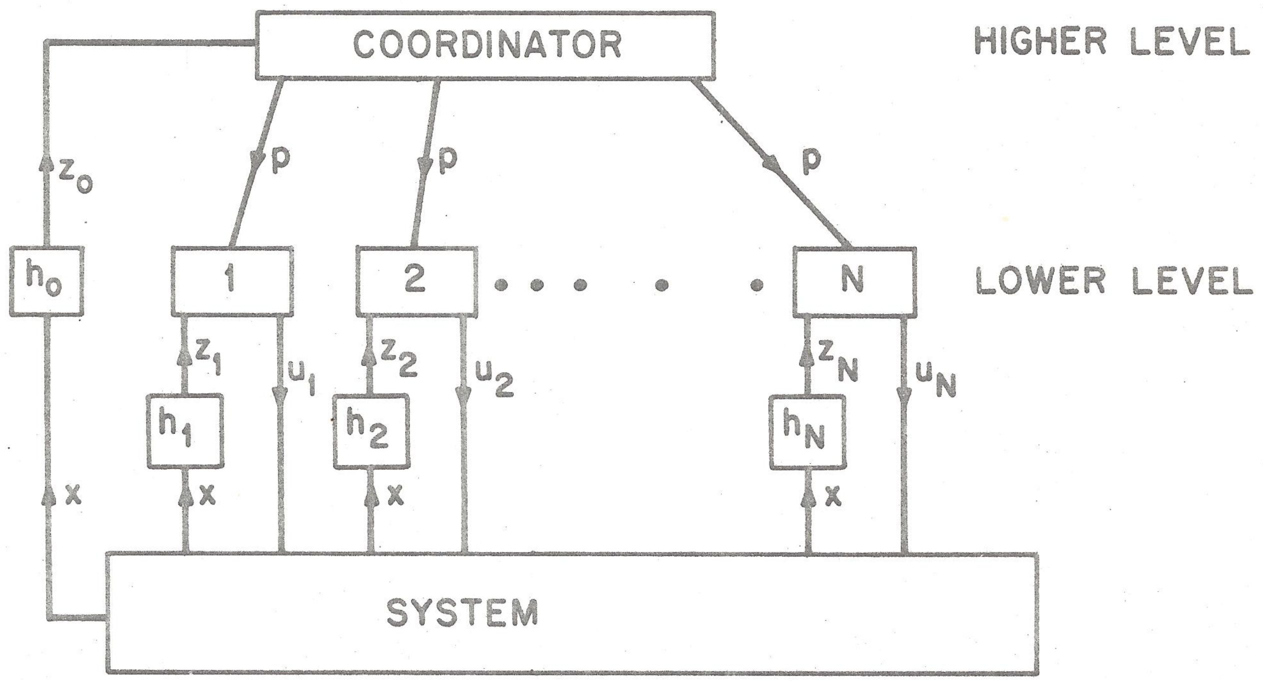


Fig. 2 Structure of Decomposition (Stochastic)

Because of the static nature of the problem, the information flow between the coordinator and lower level decision makers is only one-way.

An alternative condition for Theorem 5.3 is the following.

<u>Theorem 6.2</u>: $\tilde{L}_i(\eta_i(\cdot;z_0),\underline{p}(z_0),z_0)$ can be optimized by the ith decision maker if and only if

$$E\{f_i(\eta_i(z_i;z_0),x) + \underline{p}'(z_0)g_i(\eta_i(z_i;z_0),x)|z_i,z_0\}$$

$$= E\{f_i(\eta_i(z_i;z_0),x) + \underline{p}'(z_0)g_i(\eta_i(z_i;z_0),x)|z_i\} \qquad (6.5)$$

<u>Proof</u>: By the nested property of the conditional expectation (Appendix A)

$$\tilde{L}_i(\eta_i(\cdot;z_0),p(z_0),z_0)$$

$$= E\{E\{f_i(\eta_i(z_i;z_0),x) + \underline{p}'(z_0)g_i(\eta_i(z_i;z_0),x)|z_i,z_0\}|z_0\} \qquad (6.6)$$

If the inner conditional expectation is equal to the right side of equation (6.5), then it can be minimized with respect to $\eta_i(\cdot;z_0)$. If equation (6.5) does not hold, then $\tilde{L}_i(\eta_i(\cdot;z_0),\underline{p}(z_0),z_0)$ depends on the specific value of z_0 and thus cannot be minimized with respect to the function $\eta_i(\cdot;\underline{p}(z_0))$. Q.E.D.

We now give the results relating to the information between z_0 and z_i.

(1) $z_0 \subset z_i$. (Coordinator has less information than ith decision agent)

Then condition (6.5) is automatically satisfied.

Thus $u_i{}^* = \eta_i{}^*(z_i;\underline{p}^*(z_0))$ (6.7)

(2) $z_0 \not\subset z_i$. (Coordinator has some information not available to ith decision agent).

(a) Condition (6.5) is satisfied $u_i^* = \eta_i^*(z_i;\underline{p}^*(z_0))$

Examples: (i) $f_i(x) = f_i(x_i) \quad g_i(x) = g_i(x_i)$ (6.8)

$$z_i = h_i(x_i) \qquad z_0 = h_0([x_i]) \tag{6.9}$$

where x_i and $[x_i]$ are statistically independent.

(ii) $f_i(x) = f_i(x_i) \quad g_i(x) = g_i(x_i)$ (6.10)

$$z_i = h_i(x_i) \qquad z_0 = \begin{bmatrix} h_0^1([x_i]) \\ h_0^2(x_i) \end{bmatrix} = \begin{bmatrix} z_0^1 \\ z_0^2 \end{bmatrix} \tag{6.11}$$

$$z_0^2 \subset z_i \tag{6.12}$$

(b) Condition (6.5) is violated.

$$u_i^* = \eta_i^*(z_i;z_0)$$

$$= \eta_i^*(z_i;P(x|z_0)) \tag{6.13}$$

where $P(x|z_0)$ is the conditional probability density of x given z_0. In this case z_i and $\underline{p}^*(z_0)$ are no longer a sufficient statistics for the ith decision maker.

In words, if the coordinator has less information than the ith decision agent, as in the case when the information of the coordinator is shared by all decision agents, then the lower level problem is well defined given $\underline{p}(z_0)$ and the information of the ith decision agent. When this is not true, then the structure of the system and the information pattern has to be compatible in a certain sense, e.g. the state of the ith subsystem is statistically independent from the rest of the system and the coordinator observes that state but this information is available to the ith decision agent.

Under other circumstances, the optimization problem for the ith decision agent may not be well-defined without the knowledge of z_0.

7. Solution of the Example

Using the results derived in the previous sections, the resource manager would charge an optimal price $\underline{p}^*(z_0)$ for the resources. Each division manager would then solve the following problem.

$$\underset{\underline{\eta}_i(\cdot;\underline{z}_0)}{\text{Minimize}}\; E\{\underline{\eta}_i'(\underline{z}_i;\underline{z}_0)\underline{R}_i\underline{\eta}_i(\underline{z}_i;\underline{z}_0) - 2\underline{\eta}_i'(\underline{z}_i;\underline{z}_0)\underline{\pi}_i + \underline{p}^{*\prime}(\underline{z}_0)\underline{A}_i\underline{\eta}_i(\underline{z}_i;\underline{z}_0)|\,\underline{z}_0\} \qquad (7.1)$$

Since $\underline{\pi}_i$ is statistically independent of $\underline{v}$ and $\underline{\theta}_0$, the conditional expectation is equal to the unconditional expection given $\underline{p}^*(z_0)$. In fact the optimal $\underline{\eta}_i^*(\cdot;\underline{z}_0)$ is given by

$$\underline{\eta}_i^*(\underline{z}_i;\underline{z}_0) = \underline{R}_i^{-1}\;[E\{\underline{\pi}_i|\underline{z}_i\} - \tfrac{1}{2}\underline{A}_i'\underline{p}^*(\underline{z}_0)\}] \qquad (7.2)$$

The higher level problem is

$$\underset{\underline{p}(\underline{z}_0)\geq 0}{\text{Maximize}}\; \sum_{i=1}^{N} E\{\underline{\eta}_i^{*\prime}(\underline{z}_i^*;\underline{z}_0)\underline{R}_i\underline{\eta}_i^*(\underline{z}_i;\underline{z}_0) - 2\underline{\eta}_i^{*\prime}(\underline{z}_i;\underline{z}_0)\underline{\pi}_i + \underline{p}'(\underline{z}_0)\underline{A}_i\underline{\eta}_i^*(\underline{z}_i;\underline{z}_0)|\underline{z}_0\}$$
$$-E\{\underline{p}'(\underline{z}_0)\underline{v}|\underline{z}_0\} \qquad (7.3)$$

$$E\{\underline{\eta}_i^{*\prime}(\underline{z}_i;\underline{z}_0)\underline{R}_i\underline{\eta}_i^*(\underline{z}_i;\underline{z}_0) - 2\underline{\eta}_i^{*\prime}(\underline{z}_i;\underline{z}_0)\underline{\pi}_i + \underline{p}'(\underline{z}_0)\underline{A}_i\underline{\eta}_i^*(\underline{z}_i;\underline{z}_0)|\underline{z}_0\}$$

$$= E\{-\underline{\eta}_i^{*\prime}(\underline{z}_i;\underline{z}_0)\underline{R}_i\underline{\eta}_i^*(\underline{z}_i;\underline{z}_0)|\underline{z}_0\}$$

$$= -E\{(E\{\underline{\pi}_i|\underline{z}_i\} - \tfrac{1}{2}\underline{A}_i'\underline{p}(\underline{z}_0))'\underline{R}_i^{-1}(E\{\underline{\pi}_i|\underline{z}_i\} - \tfrac{1}{2}\underline{A}_i'\,\underline{p}(\underline{z}_0))|\underline{z}_0\}$$

$$= \tfrac{1}{4}\underline{p}'(\underline{z}_0)\underline{A}_i\underline{R}_i^{-1}\underline{A}_i'\underline{p}(\underline{z}_0) + \underline{p}'(\underline{z}_0)\underline{A}_i\underline{R}_i^{-1}\overline{\underline{\pi}}_i - c_i \qquad (7.4)$$

$$c_i = -E\{E\{\underline{\pi}_i'|\underline{z}_i\}\underline{R}_i^{-1}E\{\underline{\pi}_i|\underline{z}_i\}\} = \text{constant} \qquad (7.5)$$

Thus

$$\underline{p}^*(\underline{z}_0) = \operatorname*{Arg\ Max}_{\underline{p}(z_0) \geq 0} \; -\frac{1}{4}\underline{p}'(\underline{z}_0)(\sum_{i=1}^{N} \underline{A}_i \underline{R}_i^{-1} \underline{A}_i')p(\underline{z}_0)$$

$$+ \underline{p}'(\underline{z}_0)(\sum_{i=1}^{N} \underline{A}_i R_i^{-1} \bar{\underline{\pi}}_i - E\{\underline{v}|\underline{z}_0\}) - \sum_{i=1}^{N} c_i \qquad (7.6)$$

Comparing with the deterministic case in Section 3 we see that some kind of certainty equivalence (separation) theorem holds. The lower level devision managers choose their optimal productions by replacing the actual prices of their products with the best estimates given their measurements. However, whereas in the deterministic case the resource manager needs both $\underline{\pi}_i$, i=1,..., N and $\underline{v}$ to arrive at the optimal decision, resulting in essentially no decentralization in information, now it is only necessary to have information on $\underline{v}$.

8. Discussion and Perspectives

The decomposition achieved in mathematical programming for a class of systems with the general structure described in Section 3 is really with respect to computation. To study a possible decentralization in information we have formulated the stochastic version. It is found that under certain conditions a hierarchical decomposition for the problem is possible. The lower level decision makers need only to get certain signals from the higher level coordinator in addition to their information on the system. When these conditions are not satisfied, then in general the signals are not sufficient.

Radner and Groves [3, 9] have considered a resource allocation problem similar to the one mentioned here. However, in their treatment there exists a resource manager who is in charge of allocating the resources directly. In our formulation, the resource manager serves only a coordinator. In the deterministic case, these two formulations become the same since the lack of an information pattern reduces the problem to the case of a single decision maker.

REFERENCES

1 A. Charnes and W.W. Cooper, "Chance Constrained Programming", Management Science 6, 1959, 73-79.

2 J.L. Doob, Stochastic Processes, Wiley, New York, 1953.

3 T. Groves and R. Radner, "Allocation of Resources in a Team", Report No. 7105, Social Systems Research Institute, University of Wisconsin (Madison), April 1971.

4 P.R. Halmos and L.J. Savage, "Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics", Annals of Mathematical Statistics 20, 1949, 225-241.

5 L.S. Lasdon, "Duality and Decomposition in Mathematical Programming", IEEE Trans. Systems Sci. Cybernetics SSC-4,2, 1968, 86-100.

6 L.S. Lasdon, Optimization Theory for Large Systems, MacMillan, New York 1970.

7 D.G. Luenberger, Optimization by Vector Space Methods, Wiley, New York 1969.

8 R. Radner, "Team Decision Problems", Annals of Mathematical Statistics 33, 1962, 857-881.

9 R. Radner, "Allocation of a Scarce Resource under Uncertainty", in C.B. McGuire and R. Radner (eds.): Decision and Organization, North Holland, Amsterdam 1972.

10 J. von Neumann and O. Morgenstern, Theory of Games and Economic Behavior, Princeton Univ. Press, Princeton 1953.

APPENDIX A

SOME RESULTS IN PROBABILITY THEORY

In this appendix we summarize some definitions and results in probability theory which have been used in this paper. The probability space under consideration is denoted by $(X, \mathcal{B}, \mu)$. $\mathcal{F}$ and $\mathcal{F}_0$ are sub-σ-fields of $\mathcal{B}$.

<u>Def. A.1</u>: $\mathcal{F} \cap \mathcal{F}_0$ is the smallest-σ-field generated by $A \cap B$, where $A \in \mathcal{F}$ and $B \in \mathcal{F}_0$.

<u>Lemma A.1</u>: For any random variable ℓ, if $E\{\ell|\mathcal{F}_0\}$ is measurable with respect to $\mathcal{G}$, $\mathcal{G} \subset \mathcal{F}_0$, then $E\{\ell|\mathcal{F}_0\} = E\{\ell|\mathcal{G}\}$ a.e.

<u>Proof</u>: Given any random variable ℓ and a σ-subfield $\mathcal{G}$, the conditional expectation $E\{\ell|\mathcal{G}\}$ is characterized by two conditions:

(a) It is measurable with respect to $\mathcal{G}$;

(b) $\int_A E\{\ell|\mathcal{G}\} d\mu = \int_A \ell \, d\mu$

$$\text{for every } A \in \mathcal{G} \tag{A.1}$$

$E\{\ell|\mathcal{F}_0\}$ is measurable with respect to $\mathcal{G}$. Moreover,

$$\int_B E\{\ell|\mathcal{F}_0\} d\mu = \int_B \ell \, d\mu.$$

$$\text{for every } B \in \mathcal{F}_0 \tag{A.2}$$

Since $\mathcal{G} \subset \mathcal{F}_0$, (A.2) is also true for every $B \in \mathcal{G}$

Thus $E\{\ell|\mathcal{F}_0\}$ satisfies equation (A.1), and $E\{\ell|\mathcal{F}_0\} = E\{\ell|\mathcal{G}\}$ a.e.

Q.E.D.

<u>Lemma A.2</u>: Let γ be a $\mathcal{F} \cap \mathcal{F}_0$ - measurable function from X into U.

Let f be a measurable real-valued function on U x X. Then given any y ε X, there exists a function $\gamma(.;y)$ measurable with respect to $\mathcal{F}$ such that

$$E\{f(\gamma(x),x)|\mathcal{F}_0\}(y) = E\{f(\gamma(x;y),x)|\mathcal{F}_0\}(y) \quad \text{a.e.} \tag{A.3}$$

<u>Proof</u>: We assume two conditions, which, for this paper, will be satisfied.

(1) There exists a regular conditional probability measure $P_y^\circ(A)$.

(2) $\mathcal{F}$ and $\mathcal{F}_0$ are fields generated by functions h and h_0 so that γ being $\mathcal{F} \cap \mathcal{F}_0$ - measurable is equivalent to

$$\gamma(x) = \eta(h(x),\ h_0(x)) \tag{A.4}$$

where $\quad \eta$ is $\mathcal{A} \times \mathcal{A}_0$ - measurable on $Z \times Z_0$

$h : X \to Z$

$h_0 : X \to Z_0$

$\mathcal{A}$ and $\mathcal{A}_0$ are σ-fields on Z and Z_0

Let $\quad \gamma(x;y) = \eta(h(x),h_0(y))$. Then given y, $\gamma(.;y)$ is $\mathcal{F}$ - measurable.

$$\begin{aligned} E\{f(\gamma(x),x)|\mathcal{F}_0\}(y) &= \int_X f(\eta(h(x),h_0(x)),x)d\,P_y^\circ(x) \\ &= \int_A f(\eta(h(x),h_0(x)),x)d\,P_y^\circ(x) \\ &\quad + \int_{X-A} f(y(h(x),h_0(x)),x)d\,P_y^\circ(x) \end{aligned} \tag{A.5}$$

where

$$A = \{x;h_0(x) = h_0(y)\}\ \varepsilon\ \mathcal{F}_0 \tag{A.6}$$

Given $A \varepsilon \mathcal{F}_0$, for all $B \varepsilon \mathcal{F}_0$ (see Ref. [2])

$$\int_B P_y^\circ (A) d\, \mu(y) = \mu(A \cap B) = \int_B 1_A(x) d\, \mu(x) \qquad (A.7)$$

Therefore for all $A \varepsilon \mathcal{F}_0$,

$$P_y^\circ(A) = 1_A(y) \quad \text{for almost all } y \qquad (A.8)$$

where 1_A is the indicator funtion of A.

From equation (A.6), $y \varepsilon A$. Thus

$$P_y^\circ(A) = 1 \quad \text{for almost all } y \qquad (A.9)$$

Equation (A.5) then becomes

$$\begin{aligned} E\{f(\gamma(x),x)|\mathcal{F}_0\}(y) &= \int_A f(\eta(h(x),h_0(x)),x) d\, P_y^\circ(x) \\ &= \int_X f(\eta(h(x),h_0(y)),x) d\, P_y^\circ(x) \\ &= E\{f(\gamma(x;y),x)|\mathcal{F}_0\}(y) \qquad (A.10) \end{aligned}$$

Q.E.D.

Remark: If $\mathcal{F} = \{X, \Phi\}$ then this result reduces to the usual identity

$$E\{f(\gamma(x),x)|\mathcal{F}_0\}(y) = E\{f(\gamma(y),x)|\mathcal{F}_0\}(y) \qquad (A.11)$$

For a discussion of substitution in conditional expectation, see [1].

Lemma A.3: Let $f(u,v,y,z,x)$ be a function such that x,y,z are random variables. Suppose it is desired to choose $u(y,z)$ and $v(y)$ such that $E\{f(u(y,z),v(y),y,z,x)\}$ is minimized.

Let $u^\circ(y,z)$, $v^\circ(y)$ be the minimum of

$$\min_{\substack{u \\ v(.)}} E\{f(u,v(y),y,z,x)|y,z\}$$

Then

$$\min_{\substack{u(.,.)\\v(.)}} E\{f(u(y,z),v(y),y,z,x)\} = E\{f(u^{\circ}(y,z),v^{\circ}(y),y,z,x)\}$$

$$= E\{\min_{\substack{u\\v(.)}} E\{f(u,v(y),y,z,x)\,|\,y,z\}\} \qquad \text{(A.12)}$$

Proof:

$$E\{f(u^{\circ}(y,z),v^{\circ}(y),y,z,x)\,|\,y,z\} \leq E\{f(u(y,z),v(y),y,z,x)\,|\,y,z\}$$

for all u(.,.),v(.)

(A.13)

Thus

$$E\{f(u^{\circ}(y,z),v^{\circ}(y),y,z,x)\} = E\{E\{f(u^{\circ}(y,z),v^{\circ}(y),y,z,x)\,|\,y,z\}\}$$

$$\leq E\{f(u(y,z),v(y),y,z,x)\} \text{ for all } u(.,.),v(.)$$

(A.14)

or

$$E\{f(u^{\circ}(y,z),v^{\circ}(y),y,z,x)\} \leq \min_{\substack{u(.,.)\\v(.)}} E\{f(u(y,z),v(y),y,z,x)\}$$

(A.15)

But

$$\min_{u(.,.)} E\{f(u(y,z),v(y),y,z,x)\} \leq E\{f(u^{\circ}(y,z),v^{\circ}(y),y,z,x)\}$$

(A.16)

Hence we obtain equation (A.12)

Q.E.D.

REFERENCES (Appendix)

1 R.R. Bahadur and P.J. Bickel, "Substitution in Conditional Expectation", Annals of Mathematical Statistics 39, 1968, 377-378.

2 M. Loève, Probability Theory, Van Nostrand, Princeton 1961.
