1978

ESCORT: Engineering Systems Classification and Ordering Technique

B. A. Dendrou

S. A. Dendrou

Elias N. Houstis
Purdue University, enh@cs.purdue.edu

T. Papatheodorou

Report Number:

78-273

ESCORT: ENGINEERING SYSTEMS CLASSIFICATION AND ORDERING TECHNIQUE

B. A. Dendrou*, S. A. Dendrou*
E. N. Houstis**, T. Papatheodorou**

CSD-TR 273
July 1978

*School of Civil Engineering, Purdue University, W. Lafayette, IN 47907
**Department of Computer Science, Purdue University, W. Lafayette, IN 47907

ESCORT: Engineering Systems Classification and Ordering Technique

B. A. Dendrou*, S. A. Dendrou*,
E. N. Houstis**, T. Papatheodorou**

Abstract

Out of the study of the general problem of optimization in the presence of many objectives, there is growing a new approach which consists of ranking in order of attractiveness a discrete set of alternatives on the basis of their performance on a discrete set of specified criteria. There have been two principal lines of investigation of ordering methodologies. The one consists of proceeding by successive dichotomies of the set of alternatives until the irreducible subset (core) of equally attractive alternatives is isolated. The other consists of seeking complete (linear) orderings of the alternatives.

This report tries to pursue the second line of investigation. A methodology is proposed and implemented in the algorithm ESCORT (Engineering Systems Classification and Ordering Technique), based on the theory of inductive learning. It is a report intended to provide a preliminary treatment of the topic and show encouraging results of application.

---

*School of Civil Engineering, Purdue University, W. Lafayette, IN 47907

**Department of Computer Science, Purdue University, W. Lafayette, IN 47907

TABLE OF CONTENTS

page

ABSTRACT . . . .

1. Introduction 1

2. Scope. 2

3. Theoretical Background 5

3.1 Problem Formulation 5
3.2 Partial Orders. 6

4. Sorting Scheme 13

5. Implementation of Algorithm ESCORT 21

6. Numerical Application. 26

REFERENCES 29

APPENDICES 31

ESCORT: Engineering Systems Classification and Ordering Technique

1. Introduction

Planners and decision makers are increasingly confronted with the problem of the evaluation of social revenues of public investments. The intricacy of interactions between public and private sector and the presence of many intangible effects of public investments render difficult the estimation of the fuzzy notion of social utility. Traditionally, the effectiveness of different decisions was measured in monetary terms. Cost-benefit or cost-effectiveness techniques were used under the assumption that economic considerations were sufficient in aggregating all social goals and objectives. Today, the increasing awareness of often adverse environmental effects associated with public decisions, as well as the diversification of society's goals and objectives, bring to focus the importance of the theory of Multiple Objective Decision Making. The present work dwells with one instance of multiobjective decision making, namely the multicriteria evaluation of engineering alternatives. A short overview of the prolific pertinent literature is presented in the following section, so as to bring about the scope and aims of the present study. The theoretical aspects of the proposed model ESCORT (Engineering Systems Classification and ORdering Technique) are given in sections 3 and 4. The detail of the implementation of the algorithm is given in section 5. An example of application is given in section 6 from the area of underground engineering. Remarks on limitations and areas of potential improvement of the model conclude the report.

2. Scope

An important amount of work is reported in the recent literature on the subject of Multiple Objective Decision Making. Attention has been given to all different aspects of Decision Making. Thus some work concentrated on the analysis of the psychology of the decision making process. In this context, Game Theory [5], and the Bayesian approach [5], study the subjective aspect of decision making by dealing with the aggregation of beliefs rather than preferences. The underlying difficulty may be realized in the often encountered confusion between beliefs and actual preferences. The point of view adopted in the present work however, is on the objective side of the process. Thus, it is believed that one way of reducing arbitrariness and subjectivity from decision making in the engineering area is by improving the quality of technologic information provided by engineers in the first place. In fact, decision making can to a large extent be associated with the quest for optimization, in the larger sense of the term. However, it is recognized that for large engineering projects it is often difficult to incorporate in an optimization scheme the great amount of detail and complex mechanisms involved in physical-engineering systems. A way around this difficulty is by recognizing a certain hierarchy in the scale of the different levels of decision. Thus at the planning level of study for example, the accumulated experience and technological knowledge can be used in identifying an exhaustive list of classes of possible engineering alternatives. Such alternatives can be designed to satisfy a number of requirements such as efficiency and reliability, so as to remain comparable among each other through a set of indices of performance or attributes. A distinction is drawn at this point between the notion of attribute or characteristic and that of objective. Attributes are in this context only means towards higher ends, while the objectives or criteria provide a direct measure of the decision makers preferences. The instrumental relationship between objectives and attributes is considered here as an integral part of the multicriteria evaluation techniques. The problem then can be formulated as one of selecting the most attractive among a set of possible alternatives, on the basis of its performance on a number of criteria. This can be considered as a discrete optimization problem (selection of the optimal among a finite number of objects) as opposed to continuous optimization problems (e.g. Mathematical Programming).

Several discrete optimization techniques have been developed in recent years. They have been grouped in different categories for ease of presentation and reference. Although often partial and overlapping, such classifications do provide an insight on the merits and limitations of the different techniques and permit to detect areas of possible improvement. The classification followed here, is adapted from [7]:

According to the scope, there are techniques that aim at producing a dichotomy on a set of alternatives, one subset containing alternatives uniformly superior to the remaining ones. An example of such a technique is the multicriterion concordance method ELECTRE I [12]. Other techniques produce a cardinal evaluation of the alternatives so that an indication can be given as of how much better, given certain assumptions, one alternative is than others, all being ranked in a complete ordering. The Utility Function theory provides examples in this category, [5]. Also ELECTRE II, [4], was written for this purpose. The model ESCORT proposed here falls in the latter category.

Another classification viewpoint is based on the degree of knowledge of the decision maker's preferences. With this respect, two groups of techniques are mentioned in, [3]: Techniques which rely on prior articulation of preferences, such as ELECTRE I, [10], and techniques which make use of a progressive articulation of preferences such as the STEP method, [1], iterative weighting methods, and others. The method presented here highlights the interrelationship between articulation of preferences and value of the available information. A way is shown, for incorporating the statistical aspect of the value of information, in a multicriteria evaluation algorithm (ESCORT), considered as a first step towards a complete progressive articulation of preferences technique.

A relatively similar point of view is adopted in the distinction between descriptive, predictive and normative methods. In our view, a predictive method offers additional information useful to the decision maker in his task of selecting a best alternative, while refraining from prescribing the best solution as in the case of the normative models. In this context, a normative technique would almost necessarily have to rely on a prior knowledge of preferences.

From a computational point of view, two classes of techniques can be distinguished: those that operate in the multidimensional objective space, and those that proceed by reduction to one-dimensionality by some aggregation technique. Most techniques fall into the second category, [5], [2]. However an alternate approach exists in the context of discrete optimization. It makes use of Graph Theory and consists of building outranking relationships so as to form partial weak orders more informative than the complete order of unanimity among criteria and objectives. The latter approach is used for the model ESCORT. The theoretical background of the model is presented in the following section.

3. Theoretical Background

3.1 Problem Formulation

The problem of multicriteria evaluation of engineering alternatives as introduced in the previous section can be summarized as one of ranking a discrete set of technologically feasible alternatives for a given project, in a decreasing order of attractiveness, according to a set of decision criteria. It must be emphasized that in this perspective the determination of the alternatives will necessarily have to rely on engineering-technological knowledge, often referred to as engineering experience. This however, can be a strong argument in favor of reducing the often encountered suspicion with which most new methodologies and techniques are approached by most senior professionals. Equally importantly, the selection of the most attractive alternative will have to rely on criteria whose precision of evaluation is inversely proportional to their information content. This is in line with the interest in the notions of randomness and stochasticity, witnessed in the engineering sciences recently. The model presented hereafter can then be characterized as a framework for systematically processing a large amount of often disparate information, in view of arriving at unbiased decisions.

For ease of presentation, the following notation is introduced. The set of 'n' alternatives, referred to as objects for more generality, is denoted by:

$$(A_1, A_2, \ldots, A_n) = (A) \tag{3-1}$$

The set of 'm' criteria of selection, also referred to as objectives is denoted by:

$$(p_1, p_2, \ldots, p_m) = (p) \tag{3-2}$$

An ordering γ_p among alternatives (objects), according to criterion p is defined as a mapping:

$$\gamma_p: \; (A_i) \rightarrow K_p \qquad p = 1, \ldots, m \tag{3.3}$$

where K_p denotes a line ordering of alternatives A_i according to criterion p. If alternative A_ℓ is consistently better than all other alternatives, according to all criteria of selection, then A_ℓ is <u>unanimously</u> considered the best alternative. However, very seldom is this the case. Rather, the 'm' criteria of selection produce 'm' <u>incompatible</u> complete orderings of the 'n' alternatives. Moreover, often times the different criteria of selection are uncommensurable if not incomparable. Two complementary problems can thus be encountered. One is of selecting the "best" alternative on the basis of the above conflicting information. The other is of producing a finer overall ranking of the top alternatives so that the introduction of new criteria will permit the ultimate selection of the best alternative. The solution to either of these problems will have invariably to proceed through the determination of a <u>partial order</u> based on the 'm' complete orders, weaker than the <u>unanimity</u> condition. In this context, and in view of the above remark on the importance of the notions of randomness and stochasticity today, asserting that alternative 'i' is superior to alternative 'j' can best be considered by testing the robustness of the hypothesis that i is better than j, in the statistical sense of the term. This is the point of view adopted in the model ESCORT.

3.2 <u>Partial Orders</u>

The purpose of developing partial orders as introduced above, is to produce orderings stronger than the <u>product</u> of the 'm' complete linear orders associated with the 'm' criteria, which was referred to as the unanimity condition. Several partial orders have been reported in the pertinent literature, [4]. A well documented example is the one of the outranking relations associated with the model ELECTRE I, [12]. However, the outranking relation as defined in [10], is better used in operating a dichotomy on the set of alternatives (A_j), the one subset called the "core" or "kernel" containing a small number of incomparable among themselves but uniformly superior alternatives to the remaining subset of rejected alternatives. The partial order used in the model ESCORT is based on a hypothesis testing analogy, and makes explicit use of the value of information and information content of the criteria of selection. Thus it is best suited for engineering applications as opposed to institutional or societal systems applications. An iterative procedure is also developed for achieving a complete although weaker ordering of the alternatives through a series of consecutive dichotomies. For ease of presentation first a graph theoretic interpretation of alternatives (A) Eq. (3-1), criteria (p) Eq. (3-2), and orderings γ_p Eq. (3-3), is given.

The complete line ordering γ_p of alternatives (A_j) according to criterion 'p' can be represented by an "oriented graph" $G_p = (A, U_p)$, Fig. 3.1, whose nodes represent the alternatives (A_j) and the arcs U_p are defined by:

$$\text{Arc } (A_i, A_j) \in U_p \quad \text{if and only if} \quad \gamma_p(A_i) \geq \gamma_p(A_j) \qquad (3\text{-}4)$$

that is arc $A_i \rightarrow A_j$ denotes that alternative A_i is superior to alternative A_j. There is one such graph for each criterion ρ. These complete order graphs display the properties of transitivity and completeness. They are a consequence of the scale structure that the complete orders γ_p are assumed to have. These properties signify that there exists one arc linking any two nodes, and that the graph is circuit-free.

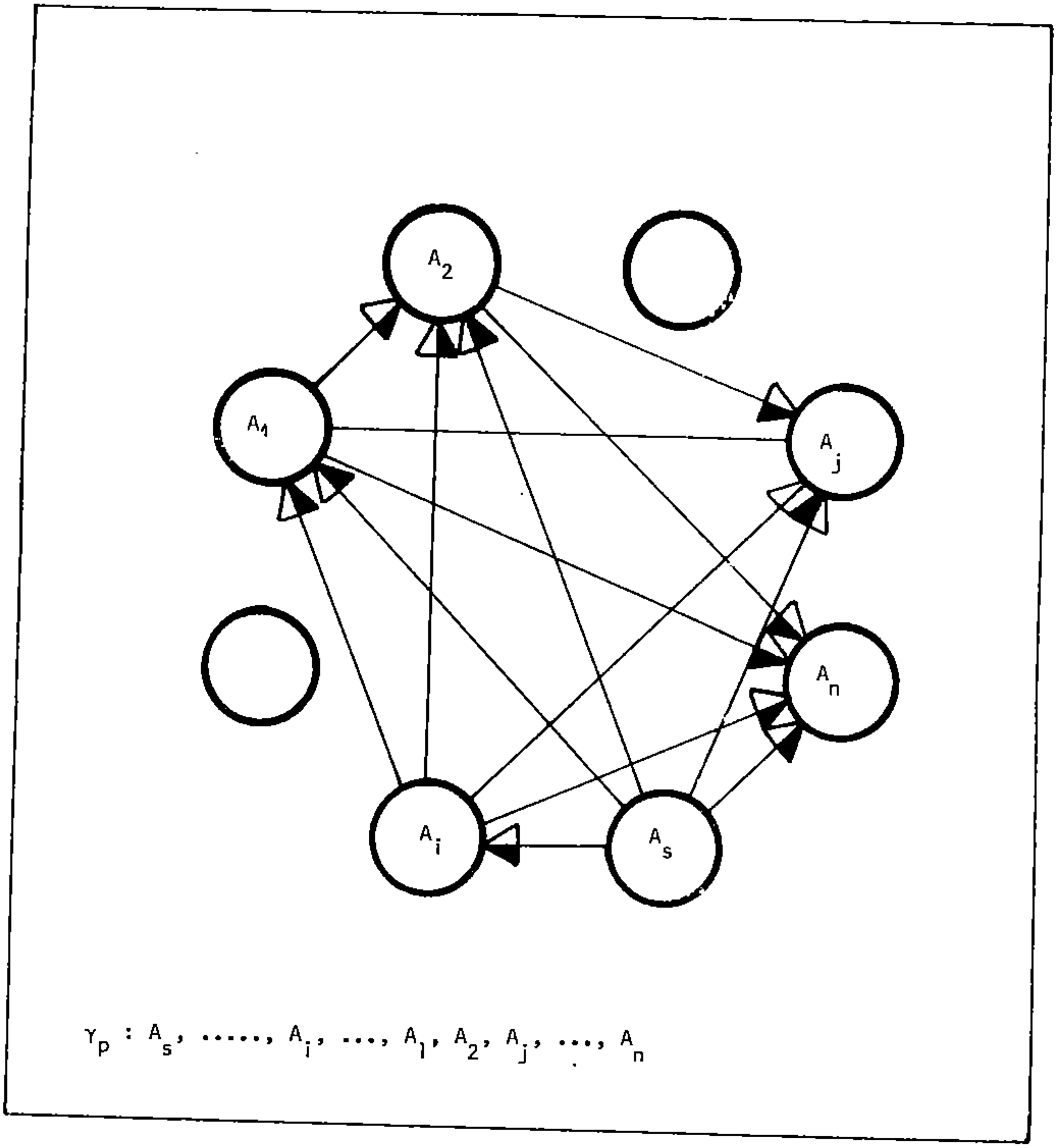


Figure 3.1 Example of a complete transitive graph $G_p = (A, \mathfrak{U}_p)$

The object of a sorting algorithm is to reduce the 'm' complete order graphs G_p to a unique graph $G = (A,U)$ that operates a synthesis of the 'm' different criteria of evaluation. This is achieved by devising partial orders

First we notice that all arcs $A_i \rightarrow A_j$ satisfying all 'm' criteria (unanimity), belong to $G(A,U)$:

$$(A_i, A_j) \in U_p \;\forall\; p = 1,\ldots m \rightarrow (A_i, A_j) \in U \tag{3-5}$$

The arcs satisfying all 'm' criteria belong to the unanimity graph

$$G_0 = (A,U_0) \tag{3-6}$$

where $$U_0 = \bigcap_{p=1}^{m} U_p$$

It can be readily seen that the unanimity graph G_0 is a subset of all possible $G = (A,U)$ graphs, containing an extremely small number of alternatives. Often times there is no alternative satisfying the unanimity condition. This serious limitation is overcome by defining a <u>partial order</u> so as to relax the too stringent unanimity condition. The 'hypothesis testing' approach used in the model ESCORT is presented hereafter.

In this approach, asserting that alternative A_i is superior to alternative A_j can at best be viewed as a more or less valid hypothesis. The plausibility of the hypothesis will then have to be based on the available information, namely the performance of each alternative on the set of criteria. In standard statistical terminology then the <u>nul hypothesis</u> would correspond to the case where the assertion that A_i is superior to A_j (denoted by A_iSA_j) is erroneous. An interesting side issue with philosophical overtones is the degree of certainty with which the hypothesis can be accepted or rejected. It is believed that ideally one can reach the desired degree of certainty at a price, by increasing the number of criteria. However, the strength of the theoretical assertion that the larger the number of criteria, the higher the certainty of the decision, can be strongly reduced by the degree of uncertainty involved in every criterion.

A typical testing of hypothesis situation is thus recognized, where a decision whether A_i is superior to A_j has to be based on a finite sample space of criteria, which in turn are attached to uncertainty due to the inherent randomness and stochasticity involved in engineering processes. This latter aspect differentiates this situation from one of equal weight ballots. Instead an iterative procedure is called for, that performs a valuation of the available criteria so as to determine a subjective weight for each criterion. Applying the hypothesis testing to all pairs of alternatives iteratively will permit to achieve a complete unbiased ranking as presented below.

Termed in a different manner, the multicriteria evaluation problem can be viewed as a problem of evaluation of hypotheses in the presence of pertinent evidence, namely the score b_{ik} of each alternative A_i over the set of criteria C_k. This problem falls in the larger category of <u>inductive learning</u> situations, [14]. The Bayesian model of inductive learning is the most natural method of applying standard probability theory. In this context, a set of hypotheses

$$H = \{H_1, H_2, \ldots, H_N\} \tag{3-7}$$

is given, whose validity is tested by a series of experiments

$$E^{(\nu)} = \{D^1, \ldots, D^{(\nu)}\} \tag{3-8}$$

where $D^{(\nu)}$ is the ν^{th} outcome of the experiment. The inductive probability or credibility $q^{(\nu)}(H_j)$ of hypothesis H_j is then defined as the conditional probability of H_j given $E^{(\nu)}$:

$$q^{(\nu)}(H_j) = P(H_j | E^{(\nu)}) \tag{3-9}$$

Equation (3-9) can be derived by the use of the Bayes formula. The convergence of $q^{(\nu)}(H_j)$ as $\nu \rightarrow \infty$ is known to be notoriously slow, [13]. It can be characterized as the process of inductive learning. Human inference is believed to differ from the above Bayesian model of induction with respect to several points. Namely, the prior probability depending on all extra-evidential information can be changed even after the series of experiments has begun. Also, equally importantly, human evaluation seems to be much more sensitive to the favorable and unfavorable evidential facts than the Bayesian model indicates. Both points are used to advantage in the model ESCORT. The result of the explicit consideration of the above points is that the credibility of the right hypothesis converges much faster, at the price of an increase in the average error probability. This scheme came to be known as superinduction, [15].

However, in the problem of multicriteria evaluation, the case of sequential, repetitive experiments is not applicable. Rather, the randomness and error in judgement is introduced by the stochastic nature of many engineering criteria. Thus, at best there is a finite probability

$$q_{ik} = P(b_{ik} \geq B_{ik}) \tag{3-10}$$

that alternative A_i will score a value of $b_{ik} \geq B_{ik}$, in the appropriate scale of criterion k. Another measure of the variability of criterion k is the coefficient of variation attached to each value b_{ik}. It should be pointed out that the larger the coefficient of variation, the smaller the intrinsic value of that criterion. In this context, the <u>value of</u>

information provided by criterion k could be measured in the absolute scale of the entropy of information, [14]:

$$S_k = - \sum_i q_{ik} \cdot \log q_{ik} \quad (3\text{-}11)$$

The summation takes place over all alternatives i.

Finally, the stage achieved in the inductive process of testing the hypotheses $H_{ij} = A_i S A_j$, that A_i is superior to A_j, could be monitored through the decrease of the inductive entropy, defined by, [15]:

$$S(H_{ij}) = - \sum_j q(H_{ij}) \cdot \log q(H_{ij}) \quad (3\text{-}12)$$

where $q(H_{ij})$ is defined in a similar fashion to Eq. (3-9).

Summarizing, while the final ranking of the alternatives (A) will have to be based and depend on the given set of criteria, the ultimate goal should be to produce an "unbiased" classification of the alternatives, by assessing the relative value of each criterion in the sense of the theory of information. This is the approach used in developing the model ESCORT. Namely, information is drawn from the sample space of criteria, so as to develop partial orders of the alternatives, based on the conjectured value of the information provided by the criteria. The definition of the partial order used in the sorting scheme of the model ESCORT follows.

4. Sorting Scheme

Denoting by $b_{\ell k}$ the score of alternative ℓ with respect to criterion k, an n x m matrix B can be formed grouping all outcomes of all alternatives with respect to all criteria:

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1k} & \cdots & b_{1m} \\ \vdots & & & & \\ b_{\ell 1} & \cdots & b_{\ell k} & \cdots & \vdots \\ \vdots & & & & \\ b_{n1} & & & \cdots & b_{nm} \end{bmatrix} = (b_{\ell k}) \qquad (4\text{-}1)$$

where the rows denote the n alternatives, and the columns the m criteria. As seen in Eq. (3-10), the scores $b_{\ell k}$ can best be considered as random variables with probabilities $q_{\ell k}$ associated with realizations $b_{\ell k} \geq B_{\ell k}$. The probabilities $q_{\ell k} = P[b_{\ell k} \geq B_{\ell k}]$, form also an n x m matrix Q

$$Q = (q_{\ell k}) \qquad (4\text{-}2)$$

The assumption is made in Eq. (3-10), without loss of generality, that larger values of $b_{\ell k}$ correspond to a better performance of alternative ℓ.

A cardinal ordering γ_k, Eq. (3-3):

$$\gamma_k : \ (A_\ell) \rightarrow K_\ell \qquad (4\text{-}3)$$

of all alternatives ℓ according to criterion k, on the basis of one set of realizations $B_{\ell k}$, is thus seen to have a finite probability of occurrence. Moreover, the aim of the study is at arriving at a complete ordering of the alternatives, on the basis of all the criteria. First, the comparison of pairs of alternatives A_i and A_j will be studied, taking into consideration all the criteria.

In line with the above discussion, the truthfulness of the assertion that alternative A_i is superior to A_j, A_iSA_j, can be measured by the probability P_{ij}:

$$P_{ij} = P[A_i S A_j] \tag{4-4}$$

The probabilities P_{ij} form an n x n matrix P whose diagonal has zeroes, since the probability that alternative A_i is superior to itself is null:

$$P = \begin{bmatrix} 0 & \cdots & P_{1i} & \cdots & P_{1j} & \cdots & P_{n1} \\ \vdots & 0 & & & & & \\ & & & & & & \vdots \\ P_{i1} & & & & P_{ij} & & \\ \vdots & & & & & & \\ P_{j1} & & P_{ji} & & & & \\ \vdots & & & & & & \\ P_{n1} & & \cdots & & & & 0 \end{bmatrix} = (P_{ij}) \tag{4-5}$$

It can be noted that the probability P_{ij}, Eq. (4-4) relates only to the pair of alternatives A_i, A_j and as such is independent of the other pairs of alternatives, and in particular the total number of alternatives. However, it depends on the criteria and more specifically on the error attached to each realization B_{ik}. Consequently, we have in general:

$$\sum_j P_{ij} \neq 1 \tag{4-6}$$

that is, matrix P, in general, is <u>not</u> a <u>stochastic</u> matrix.

The probabilities P_{ij} of Eq. (4-4) can be evaluated as follows, considering the set of criteria to be mutually exclusive:

$$P_{ij} = P[A_i S A_j]$$

$$P_{ij} = \sum_i P[A_i S A_j | \text{crit. } k] \cdot P[\text{crit. } k] \tag{4-7}$$

The first term under the summation sign can be described in plain words as the probability that $A_i S A_j$, on the basis of criterion k.

The second term under the summation sign represents the probability that criterion k is decisive in the assertion that $A_i S A_j$. It can be thought of as a weight coefficient attached to each criterion. Although a priori values can be given to these weights, it should be emphasized that they depend on the sorting process itself. A posteriori values could be used in updating iteratively the term P[crit. k] in Eq. (4-7). However, if the same weights for criterion k are used in the evaluation of all P_{ij}, the following property can be conjectured to hold:

$$P_{ij} = 1 - P_{ji} \qquad (4\text{-}8)$$

Thus, the diagonally symmetric elements of matrix P, Eq. (4-5) are seen to be complementary.

From the previous discussion on alternative scorings b_{ik} and realizations B_{ik}, Eq. (4-1), the first term under the summation sign of Eq. (4-7), can be evaluated by:

$$P[A_i S A_j | \text{crit. } k] = P[b_{ik} \geq b_{jk}] \qquad (4\text{-}9)$$

where b_{ik}, b_{jk} are random variables with probability distribution functions, respectively q_{ik}, q_{jk}, Eq. (3-10). Denoting by $f_{ij}(b_{ik}, b_{jk})$ the joint probability density function of random variables b_{ik} and b_{jk}, the r.h.s. of Eq. (4-9) can be evaluated by, (function of two random variables):

$$P[b_{ik} \geq b_{jk}] = P[b_{ik} - b_{jk} \geq 0]$$

$$P[b_{ik} \geq b_{jk}] = \int_R\int f_{ij}(b_{ik}, b_{jk})\, db_{ik}\, db_{jk} \qquad (4\text{-}10)$$

where R denotes the domain of integration $b_{ik} - b_{jk} \geq 0$, illustrated in Fig. 4.1, by the shaded area.

Noting that the alternatives A_i have been defined thus far arbitrarily, it could be plausible to assume independence among the b_{ik}, b_{jk} random variables. In this event, the joint probability density f_{ij} is equal to the product of the marginal densities:

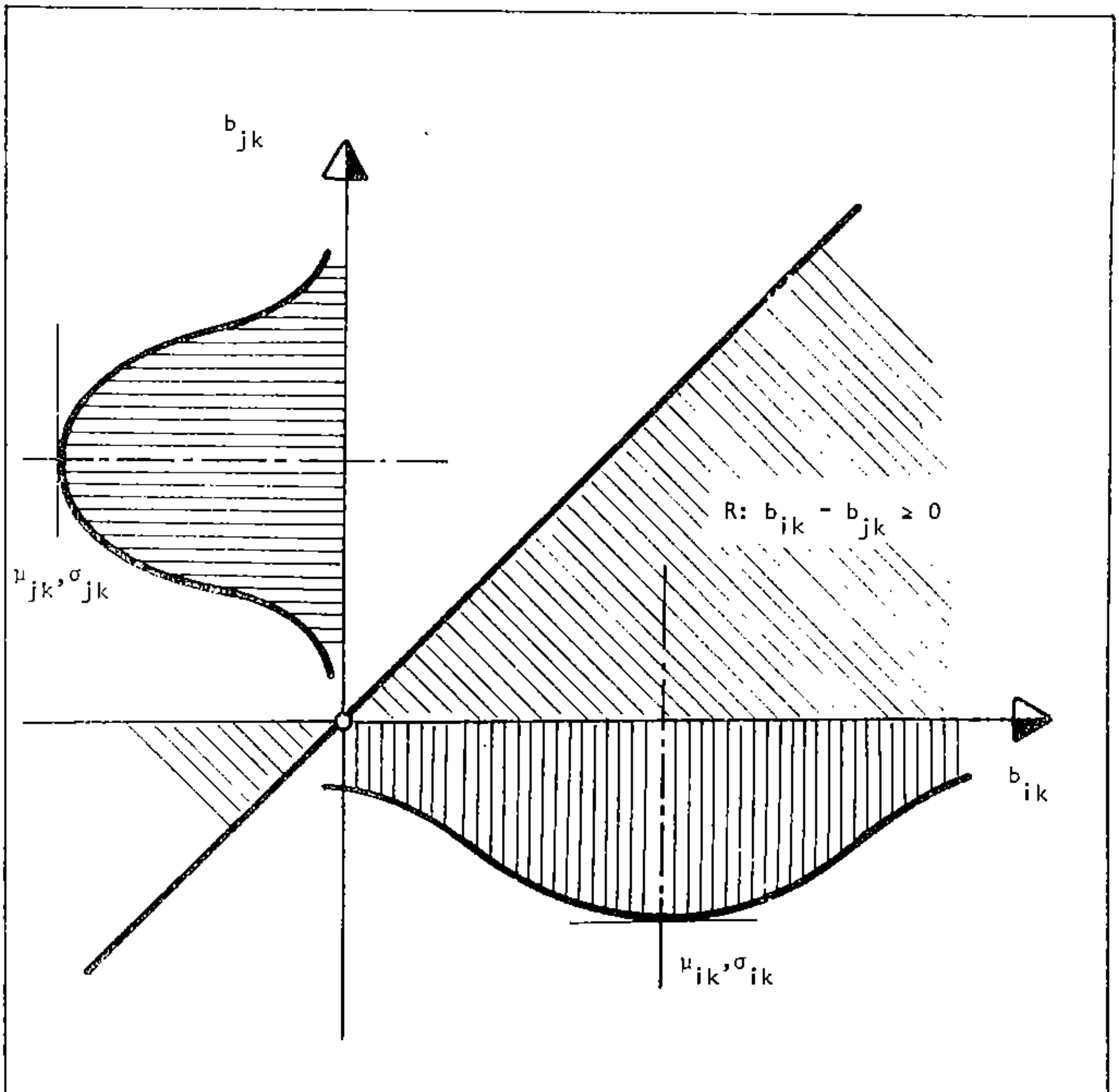


Figure 4.1. Joint Probability distribution of b_{ik}, b_{jk}, marginal densities and domain of Integration R.

$$f_{ij}(b_{ik}, b_{jk}) = f_i(b_{ik}) \cdot f_j(b_{jk}) \qquad (4\text{-}11)$$

However, if a strong correlation $r_{ij,k}$ is observed between b_{ik} and b_{jk}, the appropriately determined joint probability density should be used.

Substituting Eq. (4-9) into Eq. (4-7) we have:

$$P_{ij} = \sum_k P[b_{ik} \geq b_{jk}] \cdot P[\text{crit. } k] \qquad (4\text{-}12)$$

which is computable, making use of Eqs. (4-10) and (4-11).

Interesting information could also be provided by the probability that criterion k is decisive in the assertion that A_iSA_j, namely $P[\text{crit. } k|A_iSA_j]$. This probability can be evaluated making use of Bayes' theorem:

$$P[\text{crit. } k|A_iSA_j] = \frac{P[A_iSA_j|\text{crit. } k] \cdot P[\text{crit. } k]}{P[A_iSA_j]} \qquad (4\text{-}13)$$

Substituting Eqs. (4-9) and (4-7) into Eq. (4-13) we have:

$$F[\text{crit. } k|A_iSA_j] = \frac{P[b_{ik} \geq b_{jk}] \cdot P[\text{crit. } k]}{P_{ij}} \qquad (4\text{-}14)$$

This probability gives the <u>relative weight</u> of criterion k in the currently conjectured decision that A_i is superior to A_j, P_{ij}. It could be used to iteratively readjust the "weights" of criteria k in the evaluation of P_{ij} from Eq. (4-12), for all pairs of A_i, A_j.

Returning to the comparison of each alternative A_i with respect to all other alternatives, the following probability has to be evaluated, namely, the probability ahat alternative A_i is superior to all other alternatives, making use of the law of probabilities of joint events.

$$P[A_iSA_1 \cap A_iSA_2 \cap \ldots \cap A_iSA_{i-1} \cap A_iSA_{i+1} \cap \ldots \cap A_iSA_n] =$$
$$= P[A_iSA_1] \cdot P[A_iSA_2|A_iSA_1] \cdot P[A_iSA_3|A_iSA_1 \cap A_iSA_2] \ldots$$
$$\ldots P[A_iSA_n|A_iSA_1 \cap \ldots \cap A_iSA_{n-1}] \qquad (4\text{-}15)$$

Asserting that

$$P[A_iSA_2|A_iSA_1] = P[A_iSA_2] \tag{4-16}$$

is equivalent to assuming that the events (A_iSA_j, all i, j), are independent. Recalling the graph theoretic interpretation of section 3.2, Fig. 3.1, Eq. (4-16) is also equivalent to assuming that the resulting partial order graph will <u>not</u> have the property of <u>transitivity</u>, which implies a net gain in generality.

In this event, Eq. (4-15) reduces to

$$\Gamma_i = P[A_iSA_1 \cap \ldots \cap A_iSA_n]$$

$$\Gamma_i = \prod_{j \neq i} P[A_i \cap A_j]$$

$$\Gamma_i = \prod_{j \neq i} P_{ij} \tag{4-17}$$

Arranging the alternatives A_i in a decreasing order of the corresponding Γ_i produces the desired complete (linear) partial order. However, such a partial order would be of little use, if additional information was not given on the biasedness of the classification, its sensitivity, and the "relevance" of the provided criteria. This is attempted next.

Returning to the analogy of <u>inductive learning</u> situations mentioned in section 3.2, we are in presence of the following set of hypotheses:

$$H_i = A_iS(A_j \ ; \text{all } j \neq i) \tag{4-18}$$

that each alterrative is respectively superior to all other alternatives, or equivalently:

$$H_i = A_iSA_1 \cap A_iSA_2 \cap \ldots \cap A_iSA_{i-1} \cap A_iSA_{i+1} \cap \ldots \cap A_iSA_n \tag{4-19}$$

The validity of these hypotheses is conjectured and tested on the basis of the performance criteria. It is measured by the probability of truthfulness Γ_i, as given by Eq. (4-17).

In turn the validity of a ranking of alternatives A_i on the basis of the probabilities Γ_i, can be measured by the degree of <u>inductive entropy</u> achieved:

$$S(H_i) = - \sum_i \Gamma_i \cdot \log (\Gamma_i) \tag{4-20}$$

According to the law of entropy decrease, the smaller the entropy $S(H_i)$, the stronger the faith attributed to the ranking Γ_i. Substituting Eq. (4-17) into Eq. (4-20) we have successively the following transformations, making use of the properties of logarithms and exponentials:

$$S(H_i) = - \sum_i \left(\prod_{j \neq i} P_{ij} \right) \cdot \log \left(\prod_{j \neq i} P_{ij} \right)$$

$$= - \sum_i \log \left(\left(\prod_{j \neq i} P_{ij} \right)^{\prod_{k \neq i} P_{ik}} \right)$$

$$= - \sum_i \log \left(\prod_{j \neq i} (P_{ij})^{\prod_{k \neq i} P_{ik}} \right)$$

$$= - \sum_i \sum_{j \neq i} \log \left((P_{ij})^{\prod_{k \neq i} P_{ik}} \right)$$

$$= \sum_i \sum_{j \neq i} \left(\prod_{k \neq i} P_{ik} \right) \cdot (-P_{ij} \log P_{ij})$$

$$= \sum_i \underbrace{\left(\prod_{k \neq i} P_{ik} \right)}_{= \Gamma_i} \sum_{j \neq i} (S_{ij})$$

$$S(H_i) = \underbrace{\sum_i \Gamma_i}_{= 1} \sum_{j \neq 1} S_{ij} = \sum_i \sum_{j \neq i} S_{ij} \tag{4-21}$$

where:

$$S_{ij} = -P_{ij} \log P_{ij} \tag{4-22}$$

is the inductive entropy of hypothesis $H_{ij} = A_i S A_j$, over all criteria k. Equation (4-21) results from the assumption of independence among the set of hypotheses ($H_{ij} = A_i S A_j$, all i,j).

If the probabilities P_{ij} are determined iteratively as previously suggested, a condition for the series $S^{(\nu)}(H_i)$ defined in Eq. (4-21) to be decreasing is, that the individual inductive entropies S_{ij} decrease.

The consideration of the notion of inductive entropy in conjunction with the sensitivity of the partial orders delineated above, concludes the theoretical background necessary in the development of the sorting scheme of the model ESCORT. The implementation of the algorithm, as well as convergence considerations will be presented next.

5. Implementation of Algorithm ESCORT

The sorting or ordering scheme developed in section 4 is primordially based on a pair-wise comparison of all alternatives (A_j). This is done by iteratively determining the probabilities P_{ij} that alternative A_i is superior to alternative A_j, $A_i S A_j$, Eq. (4-4), which are grouped in matrix P, Eq. (4-5). Fundamental in the evaluation of the probabilities P_{ij} are the probabilities $P[b_{ik} \geq b_{jk}]$, Eq. (4-12), that alternative A_i scores higher (better) than alternative A_j according to criterion k. First, expressions for the evaluation of these joint-event probabilities are developed. They are generally evaluated by the r.h.s. expression of Eq. (4-10)

$$P[b_{ik} \geq b_{jk}] = \iint_R f_{ij}\ (b_{ik}, b_{jk})\ db_{ik}\ db_{jk} \qquad (5\text{-}1)$$

where f_{ij} is the joint probability density function of random variables b_{ik} and b_{jk}.

It is shown in Appendix A that on the basis of information about a random variable limited to its mean value and standard deviation, an unbiased estimate of the probability density function is the normal p.d.f. This could be the case for the random variables b_{ik}. Thus, if the random variables b_{ik} and b_{jk} have a correlation coefficient $r_{ij,k}$, their joint p.d.f. is, [8]:

$$f_{ij} = \frac{1}{2\pi\sigma_i\sigma_j\sqrt{1-r_{ij,k}^2}} \cdot$$

$$\cdot \exp\left[-\frac{1}{2(1-r_{ij,k}^2)}\left\{\frac{(b_{ik}-\mu_i)^2}{\sigma_i^2} - \frac{2r_{ij,k}(b_{ik}-\mu_i)(b_{jk}-\mu_j)}{\sigma_i \cdot \sigma_j} + \frac{(b_{jk}-\mu_j)^2}{\sigma_j^2}\right\}\right] \qquad (5\text{-}2)$$

where μ_i, μ_j and σ_i, σ_j are, respectively, the mean values and standard deviations of b_{ik} and b_{jk}.

As seen in Fig. 4.1, the domain of integration R, $b_{ik} \geq b_{jk}$ can also be viewed as $b_{ik} - b_{jk} \geq 0$. In this case, the p.d.f. of $b_{ik} - b_{jk}$ is evaluated by, [9]:

$$f_{i-j}(b_{jk}) = \int_{-\infty}^{\infty} f_{ij}(b_{ik}, b_{ik}-b_{jk})\, db_{ik} \tag{5-3}$$

and the integral of the r.h.s. of Eq. (5-1) reduces to:

$$P[b_{ik} - b_{jk} \geq 0] = 1 - P[b_{ik} - b_{jk} \leq 0]$$

$$= 1 - \int_{-\infty}^{\infty} f_{i-j}(b_{jk})\, db_{jk} \tag{5-4}$$

In the case where b_{ik} and b_{jk} are normal uncorrelated r.v. (independent r.v., $r_{ij,k} = 0$), it is shown, [9], that f_{i-j} is also normal with the following mean and variance:

$$m = \mu_i - \mu_j$$

$$\sigma^2 = \sigma_j^2 + \sigma_j^2 \tag{5-5}$$

Eq. (5-4) is thus evaluated by:

$$P[b_{ik} - b_{jk} \geq 0] = 1 - \Phi\left(\frac{0-m}{\sigma^2}\right) \tag{5-6}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function at ordinate $(\cdot)$. Eq. (5-6) is used in the example of application of the following section.

The evaluation of the pair-wise probabilities $P[b_{ik} \geq b_{jk}]$ constitutes the first step of the algorithm ESCORT as shown in the flowchart of Figure 5.1. This, along with the evaluation of the total pair-wise probabilities P_{ij} (Eq. 4-12), for the equiprobable values of $P[\text{crit } k] = 1/m$, conclude the initialization phase of the algorithm.

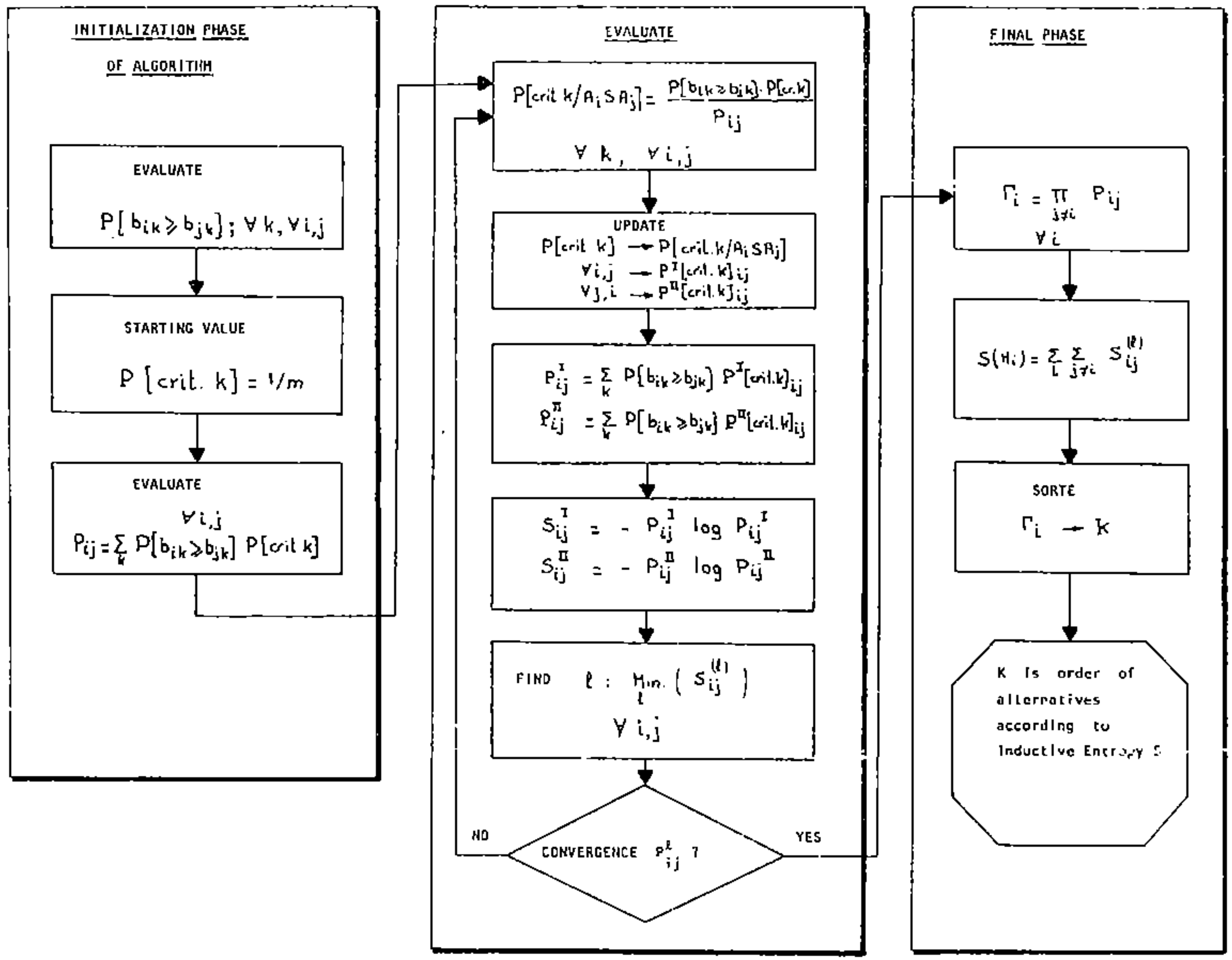


Figure 5.1 Flowchart of Algorithm ESCORT

The following step consists of evaluating the conditional probabilities of the decisiveness of criterion k in the assertion that $A_i S A_j$, by Eq. (4-14) (Bayes' theorem). It is important to notice that two sets of values are associated to every criterion k, corresponding to the complementary orders $i \to j$ and $j \to i$. They are a priori equally valid. If these two sets of values for P[crit. $k|A_i S A_j$] and P[crit. $k|A_j S A_i$] are used to replace the equiprobable values P[crit. k] in Eq. (4-12), two new sets of values for P_{ij} are produced, denoted $P_{ij}^{(I)}$ and $P_{ij}^{(II)}$, Fig. 5.1.

However, one set among the two sets of values for P_{ij} produces a net gain in information as measured by the decrease in the inductive entropy S_{ij}, Eq. (4-22). This is the set to be retained for updating the decisiveness (or weight) probabilities P[crit. $k|A_i S A_j$], and so forth iteratively until the scheme converges to stable values for the P_{ij}'s. The convergence of the scheme will be addressed later. The final values of P_{ij} can subsequently be used to evaluate the probabilities Γ_i that alternative A_i is superior to all other alternatives, by Eq. (4-17). The sorting of the Γ_i provides the complete (linear) order K of the alternatives (A_j) that was originally sought. Moreover, the corresponding value of the inductive entropy $S(H_i)$, Eq. (4-20), provides a measure of the fidelity that one can have on the order K.

Concerning the convergence of the algorithm developed above, the following sketch of a proof can be proposed. Denoting by $P_{ij}^{(\ell)}$ the value of the probability P_{ij} at the ℓ-th iteration, a convergence condition would require the absolute value of the change in the probabilities P_{ij} between successive iterations to be bound:

$$\left| P_{ij}^{(\ell)} - P_{ij}^{(\ell-1)} \right| < A \qquad (5\text{-}7)$$

Substituting $P_{ij}^{(\ell)}$ and $P_{ij}^{(\ell-1)}$ from Eqs. (4-12) and (4-14) we have:

$$P_{ij}^{(\ell)} = \sum_k P[b_{ik} \geq b_{jk}] \cdot P[\text{crit. } k|A_i SA_j]^{(\ell)} \tag{5-8}$$

$$P_{ij}^{(\ell-1)} = \sum_k P[b_{ik} \geq b_{jk}] \cdot P[\text{crit. } k|A_i SA_j]^{(\ell-1)} \tag{5-9}$$

Thus we have:

$$P_{ij}^{(\ell)} - P_{ij}^{(\ell-1)} = \sum_k P[b_{ik} \geq b_{jk}] \cdot \{P[\text{crit. } k|A_i SA_j]^{(\ell)} - P[\text{crit. } k|A_i SA_j]^{(\ell-1)}\} \tag{5-10}$$

Applying the Hoelder inequality then we have:

$$\begin{aligned} \left|P_{ij}^{(\ell)} - P_{ij}^{(\ell-1)}\right| &\leq \sum_k \left|P[b_{ik} \geq b_{jk}]\right| \cdot \left|P[\text{crit. } k|A_i SA_j]^{(\ell)} - P[\text{crit. } k|A_i SA_j]^{(\ell-1)}\right| \\ &\leq \{\sum_k P[b_{ik} - b_{jk}]\} \cdot \\ &\quad \cdot \max_k (\left|P[\text{crit. } k|A_i SA_j]^{(\ell)} - P[\text{crit. } k|A_i SA_j]^{(\ell-1)}\right| \end{aligned} \tag{5-11}$$

which is a Lipschitz condition guaranteeing convergence if:

$$\max_k \left\{\left|P[\text{crit. } k|A_i SA_j]^{(\ell)} - P[\text{crit. } k|A_i SA_j]^{(\ell-1)}\right|\right\} < 1 \tag{5-12}$$

since $\sum_k P[b_{ik} \geq b_{jk}]$ is a constant and bound quantity.

Noticing that quantities figuring in Eq. (5-12) are all probabilities, we can presume that the inequality in general holds, guaranteeing the convergence of the algorithm. In fact, the example of application of the following section showed a fast rate of convergence.

6. Numerical Application

A hypothetical example is given hereafter. It is only meant to provide a clear illustration of the implementation of the algorithm ESCORT. The treated case is inspired by an underground excavation and the main body of data concerning the physical parameters of the geologic media are taken from (B. A. Dendrou et al, "Dynamic Uncertainty Analysis," C.S. report No. 205).

Thirty-six different locations are considered for a preliminary design of a power plant which is a part of a large hydrologic system. Each location constitutes in fact, a particular alternative for the project.

From the technical point of view the performance of each alternative can be described by a number of attributes (objectives). More precisely seven attributes are retained, they are the following: The Modulus of Elasticity of the geologic media, The Poisson's ration, the initial horizontal stresses, the Dynamic displacements, the Dynamic stresses, and the piezometric head.

The mean values and the coefficients of variation of the above physical quantities are provided at each location and can be computed by an inference numerical scheme, (B. A. Dendrou et al, "A Methodology to Evaluate Static and Dynamic Design Criteria for Underground Openings," 19th Symposium on Rock Mechanics, May 1978).

Table 6.1 provides the spatial distribution of the above mentioned data values and Figures 6.1, 6.2, 6.3 and 6.4 illustrate them at the thirty-six chosen locations which lie on a square [500 x 500] and form a 6 x 6 mesh.

Table 6.1 Input Statistical values of the given attributes at different location points.

LOCATION NUMBER	MODULUS OF ELASTICITY		UNIT WEIGHT		POISSON'S RATIO		PIEZOMETRIC HEAD		INITIAL HORIZONT. STRESSES		DYNAMIC STRESSES		DYN. DISPLACEMENTS	
	MEAN	C.V.	MEAN	C.V.	MEAN	C.V.	MEAN	C.V.	MEAN	C.V.	MEAN	C.V.	MEAN	C.V.
1	28.03	0	1.	0	10.0		12.0	0.98	0.16	0.0	1.78	0.09	-8.6	0.0
2	19.74	0.29	1.09	0.29	13.76		6.4	0.96	0.15	.01	1.78	0.09	-2.55	0.1
3	14.39	0	1.	0	10.0		7.7	0.65	0.155	.02	.33	0.23	-1.05	0.09
4	10.05	0.63	1.01	0.44	11.7		8.6	0.11	0.135	.3	.088	1.01	- .22	.7
5	6.33	0.33	1.04	0.11	14.63		4.1	0.10	0.14	2.0	.02	0.03	.51	4.0
6	5.29	0	1.	0	10.0		0	0.0	0.1	.04	0.035	0.6	- .34	3.0
7	38.72	0.33	1.054	0.33	11.48		13.6	0.02	0.162	0.0	1.17	0.17	-6.74	0.0
8	27.80	0.10	1.056	0.1	26.72		11.6	0.59	0.153	.02	1.17	0.17	-1.1	0.02
9	20.32	0.138	1.05	0.13	30.43		8.9	0.14	0.14	6.0	.103	0.19	- .60	1.0
10	14.24	0.154	1.045	0.05	33.74		7.0	0.14	0.13	.5	.03	0.002	- .01	1.0
11	10.01	1.1	1.044	0.06	26.94		4.6	0.04	0.10	3.0	.012	0.3	.38	3.0
12	7.12	1.49	1.041	0.1	19.82		0.04	0.0	0.11	.05	.029	0.04	- .63	2.0
13	54.59	0	1.035	0	10.0		13.3	0.52	0.2	0.0	2.566	0.16	-7.71	0.0
14	39.39	0.16	1.091	0.47	33.77		11.5	0.13	0.17	.04	2.566	0.16	- .79	.06
15	27.22	0.26	1.085	0.10	42.92		9.8	0.20	0.14	.5	.15	0.14	- .30	3.0
16	20.13	0.04	1.078	0.23	45.97		7.8	0.23	0.14	.7	.08	0.003	.13	.7
17	14.63	0.37	1.07	0.11	34.47		4.5	0.05	0.12	.08	.03	0.03	.132	3.0
18	9.86	2.37	1.086	0.19	8.40		0.01	0.4	0.12	1.0	0.035	0.04	- .54	1.0
19	78.30	0.28	1.12	0.28	15.33		12.6	0.3	0.19	0.0	1.35	0.10	-2.93	0.0
20	53.97	0.02	1.11	0.02	32.79		12.6	0.09	0.18	.2	1.35	0.003	- .28	.4
21	38.75	0.033	1.1	0.21	44.89		10.0	0.21	0.15	.4	0.10	0.003	- .04	.9
22	27.90	0.17	1.093	0.2	40.24		8.5	0.2	0.13	.5	.04	0.02	.17	.6
23	20.51	1.6	1.085	0.11	28.08		5.0	0.93	0.10	.5	0.02	0.04	.108	2.0
24	8.71	2.3	1.037	0.65	5.26		0.0	0.0	0.18	2.0	0.034	0.31	- .22	.8
25	109.5	0.90	1.146	0.90	15.08		15.5	0.07	0.17	0.0	3.72	0.11	- .41	0.0
26	53.97	0.018	1.136	0.06	27.05		13.6	0.06	0.16	1.0	3.72	0.002	- .05	1.0
27	54.82	0.099	1.126	0.03	33.55		14.5	0.018	0.13	.3	.23	0.10	.044	.7
28	39.28	0.29	1.445	0.13	34.68		9.1	0.17	0.11	5.0	.10	0.05	.18	.6
29	28.52	0.1	1.446	0.53	26.73		5.6	0.10	0.12	.7	.03	0.31	.15	9.0
30	19.97	0.123	1.06	0.16	10.25		0.01	0.29	0.19	3.0	.041	0.31	- .039	1.0
31	148.4	0	1.165	0	10.0		16.5	0.0	0.175	0.0	2.48	0.19	0.0	0.0
32	107.3	0.069	1.155	0.07	110.7		14.6	0.9	0.26	.6	2.48	0.52	0.003	.9
33	77.59	0.24	1.126	0.3	17.28		12.2	0.28	0.14	.1	.15	0.13	.06	1.0
34	56.76	0.65	1.13	0.52	25.18		9.5	0.0	0.12	.4	.07	0.01	.18	.7
35	10.05	0.69	1.13	0.02	22.74		5.4	0.33	0.12	.3	.03	0.02	.14	2.0
36	6.93	0.33	1.12	0.3	10.0		0.0	0.0	0.12	1.0	.032	0.03	-0.014	.9

Two sets of runs are performed to assess the validity of the numerical scheme.

The first set consists of three runs in which the attributes are taken in pairs. The results are given in Table 6.2 and are conformal to the common sense. The most appealing locations are identified and are in agreement with our expectations (see Figures 6.3 and 6.4).

The second set of runs concerns all the chosen attributes considered simultaneously. The following observations are made.

a. The convergence of the scheme is obtained after eight to nine iterations (Figure 6.5);

b. The relative importance of each attribute is reflected by conditional probability as defined in Eq. (4-14) after the scheme's convergence is achieved;

c. A definite judgement on the validity of the considered set of attributes can be made through the use of the inductive entropy as given in Eq. (4-22);

d. The ranking of each alternative is characterized by the following coefficient, reflecting the weight with which the sorting process is obtained;

$$W_i = \frac{\Gamma_i}{\sum_K^n \Gamma_K} \tag{6-1}$$

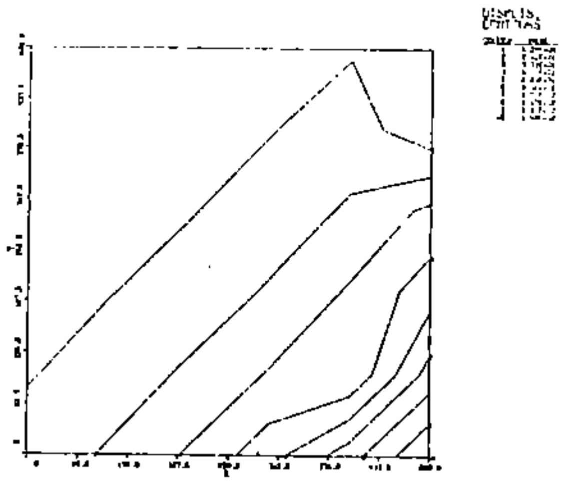
where Γ_i are the probabilities defined in Eq. (4-12).

The most appealing locations (alternatives) are depicted through the comparison of their performance criteria as suggested above. In the specific case of the treated example, location 33 is most advantageous, as anticipated from common engineering judgement. As a general rule,

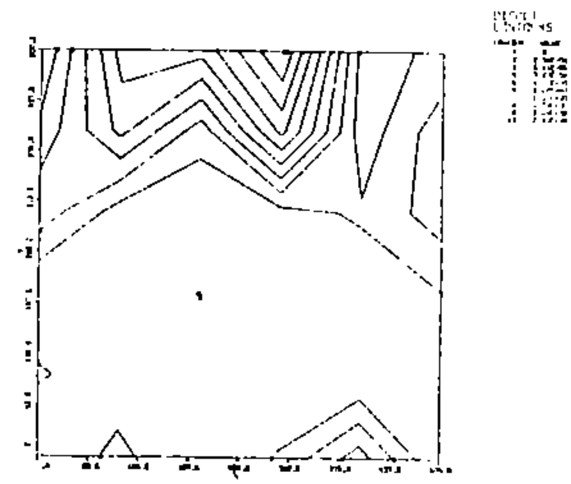
Table 6.2. Results of the Algorithm's Application

NUMBER OF THE RUN	CONSIDERED ATTRIBUTES	THREE MOST APPEALING LOCATIONS	RELATIVE WEIGHT	ENTROPY OF INFORMATION	RETAINED LOCATION (ALTERNATIVE)
601	DYN. STRESSES UNIT WEIGHT	13 31 19	.40 .27 0.10	8.17 3.93 9.14	
602	POISSON'S RATIO MODUL. ELAST.	24 36 30	.68 .31 .03	3.55 3.88 4.25	
603	PIEZOM. HEAD. HOR. STRESSES	31 12 7	.45 .38 .11	6.97 2.024 7.045	
650	DYN. DISPLAC. DYN. STRES. MOD. ELAST.	4 35 16	.65 .21 .07	5.13 5.90 6.25	
605	DYNAMIC STRESSES MODULUS OF ELAST. DYNAMIC DISPL. PIEZOM. HEAD	36 4 35	.94 .04 .005	4.59 5.33 5.99	
607	DYNAM. STRESSES MODULUS OF ELAST. DYN. DISPLAC. UNIT WEIGHT POISSON'S RATIO	33 36 34	.95 .04 .003	3.53 4.82 5.68	←

MODULUS OF ELASTICITY

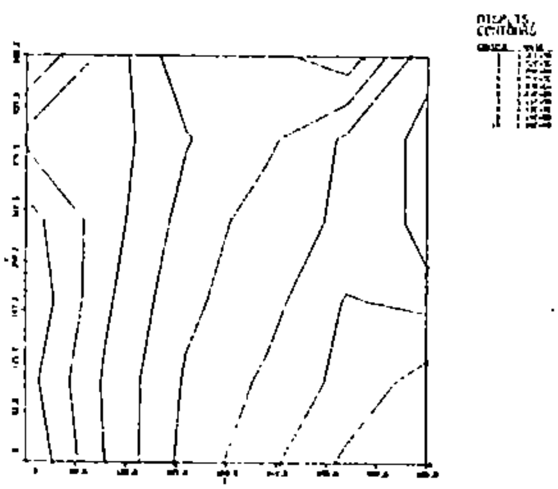


MEAN VALUE

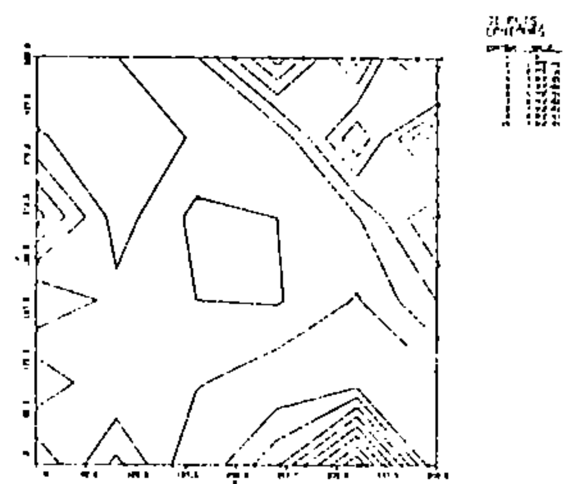


COEF. OF VARIATION

UNIT WEIGHT



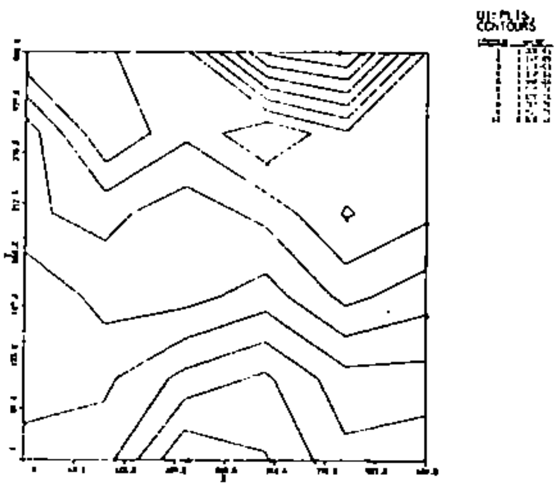
MEAN VALUE



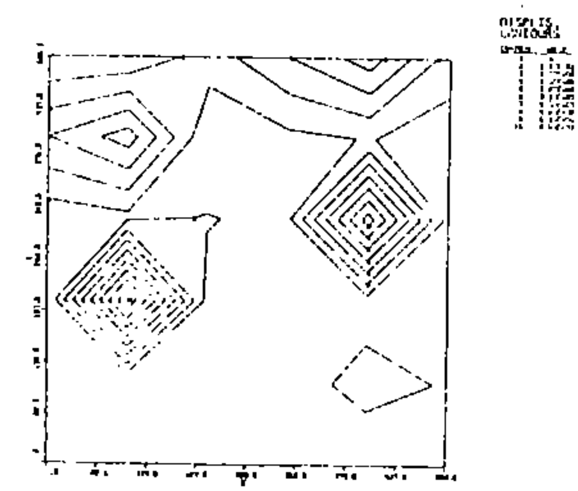
COEF. OF VARIATION

Figure 6.1 Spatial Distribution of Attributes

HORIZONTAL STRESSES

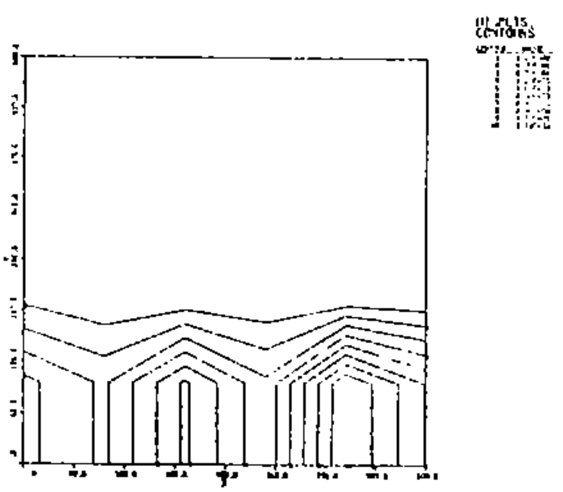


MEAN VALUE

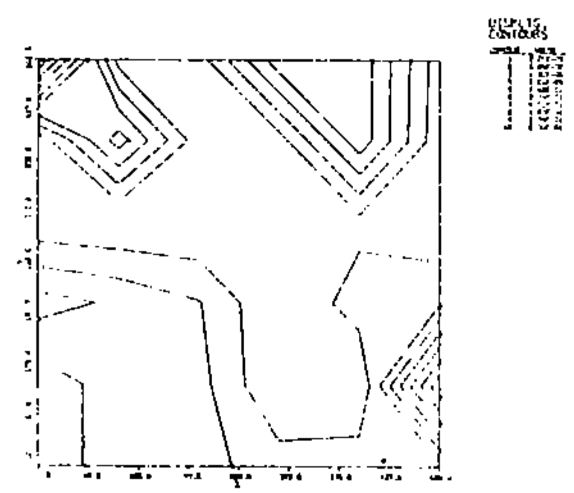


COEF. OF VARIATION

DYNAMIC STRESSES



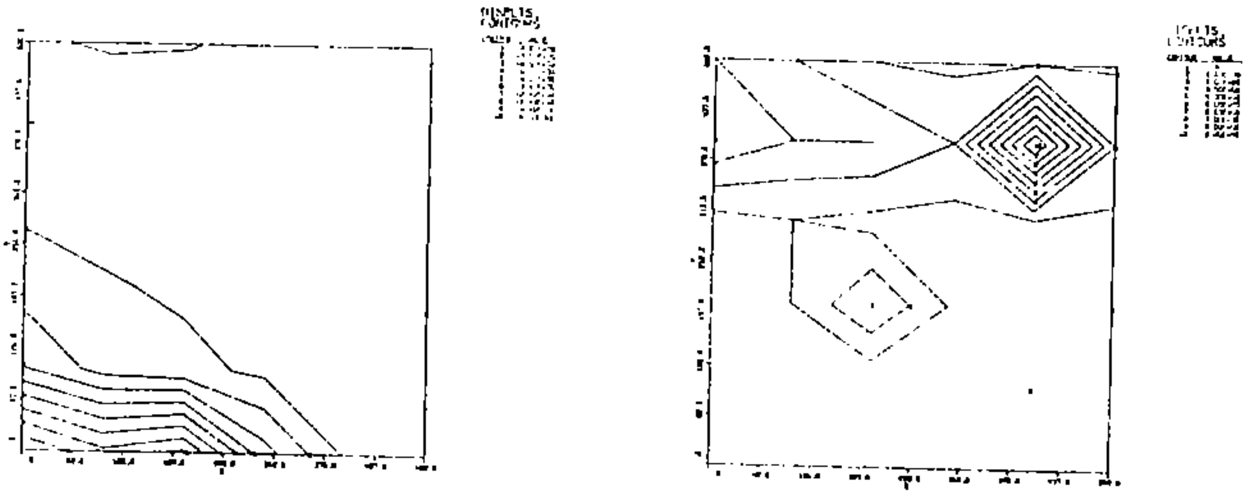
MEAN VALUE



COEF. OF VARIATION

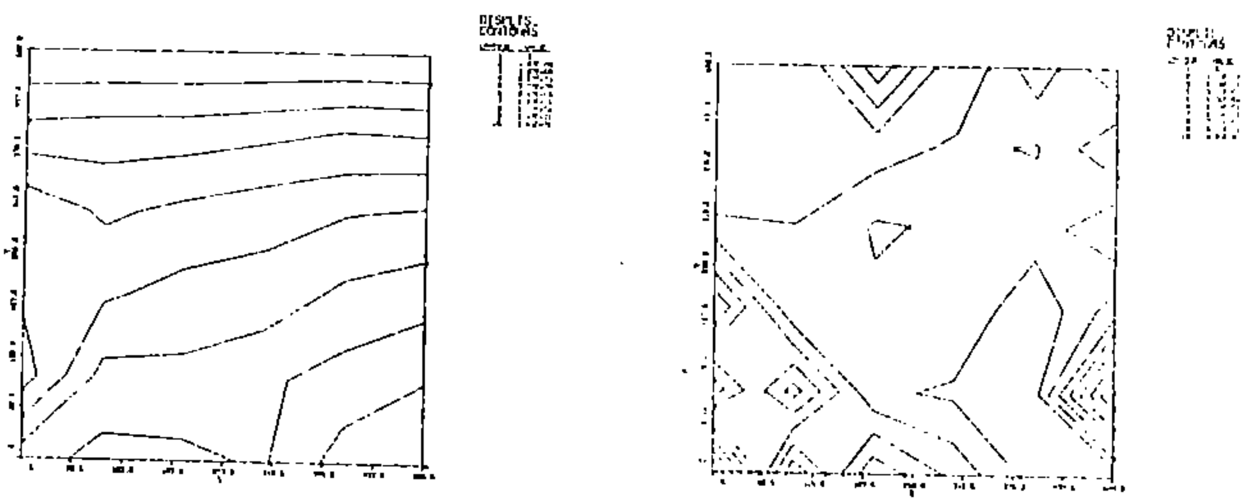
Figure 6.2. Spatial Distribution of Attributes

DYNAMIC DISPLACEMENTS



MEAN VALUE | COEF. OF VARIATION

PIEZOMETRIC HEAD



MEAN VALUE | COEF. OF VARIATION

Figure 6.3. Spatial Distribution of Attributes

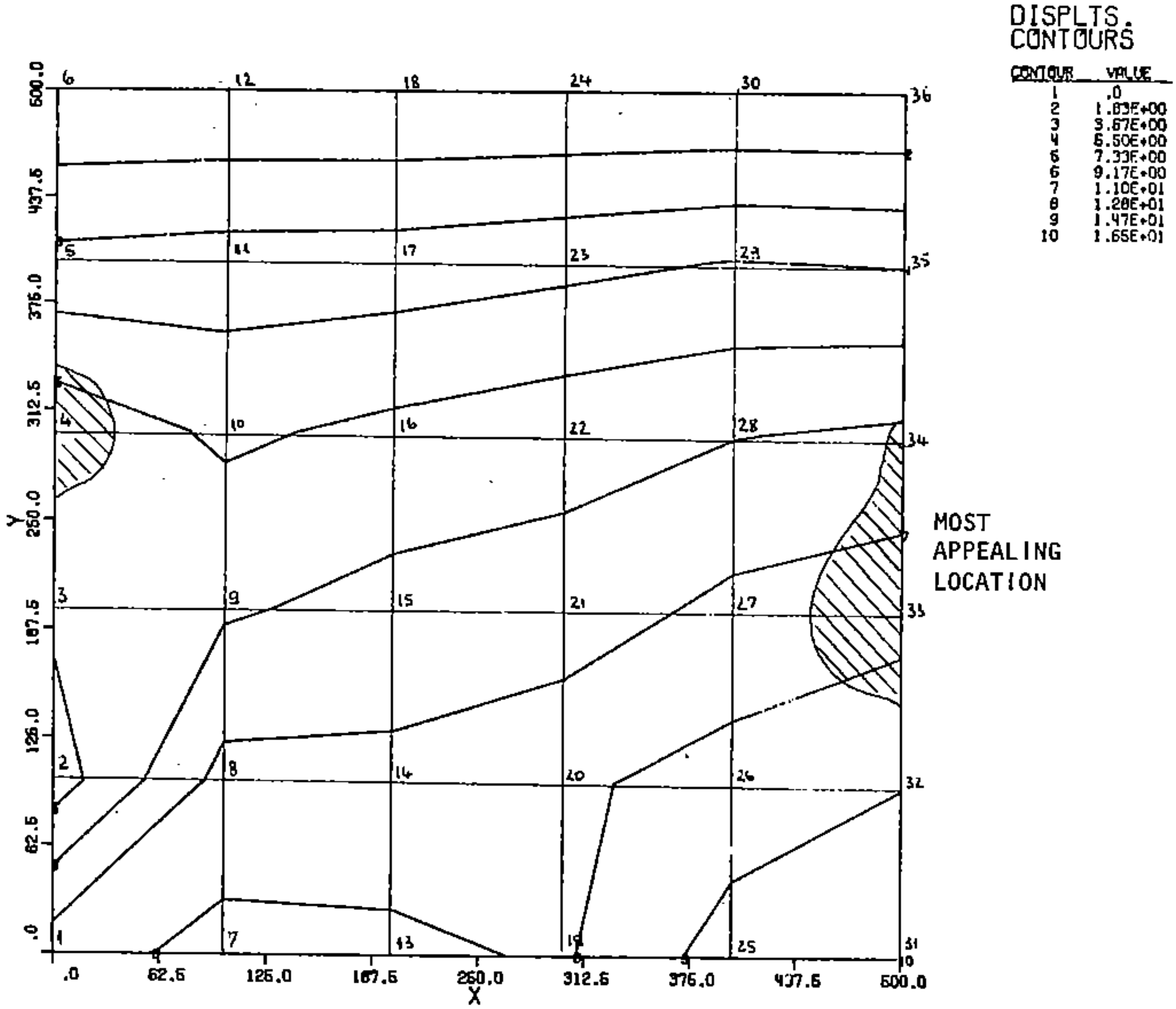


Figure 6.4. Location of Most Appealing Alternative

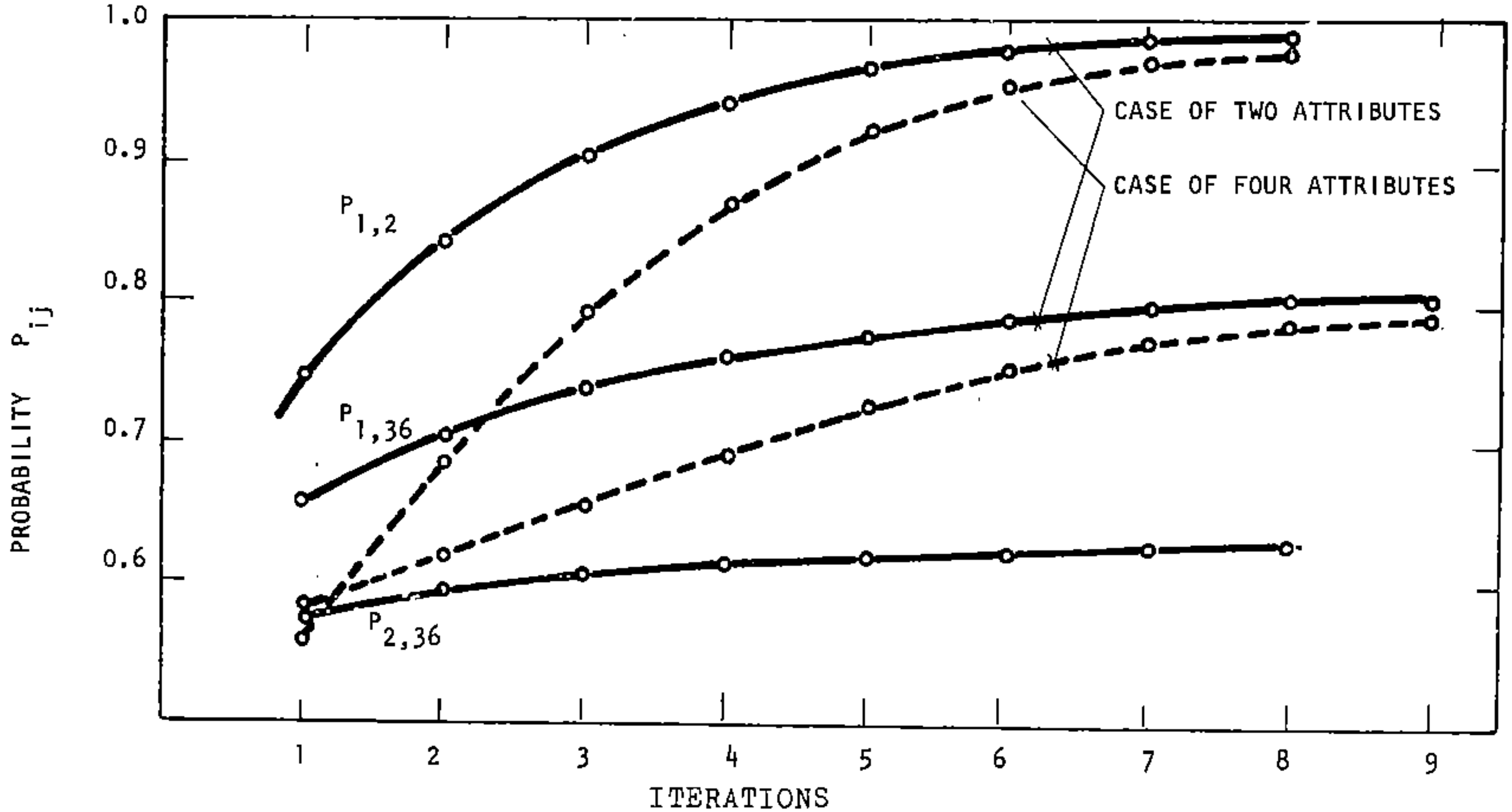


Figure 6.5 Convergence of the Algorithm

it can be conjectured that the inductive entropy decreases as the number of attributes (criteria) increases. This is realistic, since a larger number of attributes corresponds to an increase in the available information.

REFERENCES

1. P. Bertier, J. M. Bouroche, "Analyse des Données Multidimensionnelles," Presses Universitaires de France, Paris, 1975, Chapter III.

2. J. M. Blin, "The General Concept of Multidimensional Consistency: Some Algebraic Aspects of the Aggregation Problem," Selected Proceedings of a Seminar on Multiple Criteria Decision Making, edited by J. Cochrane and M. Zeleny, University of South Carolina Press, Columbia, South Carolina, 1973.

3. J. L. Cohon, D. H. Marks, "A Review and Evaluation of Multiobjective Programming Techniques," Water Resources Research, Vol. 11(2), April 1975, pp. 208-220.

4. E. Jacquet-Lagrèze, "How We Can Use the Notion of Semiorders to Build Outranking Relations in Multicriteria Decision Making," paper presented at the Fourth Research Conference on Subjective Probability, Utility and Decision-Making, Rome, September 1973, Selected Proceedings ... , D. Wendt, C. Vlek, editors, 1975, pp. 87-112.

5. R. L. Keeney, H. Raiffa, "Decisions with Multiple Objectives," John Wiley, New York, 1976.

6. V. L. Kupershtokh, B. G. Mirkin, "Ordering of Interrelated Objects, II," Automation and Remote Control, January 1972, pp. 1093-1098.

7. K. R. McCrimmon, "An Overview of Multiple Objective Decision Making," Selected Proceedings of a Seminar on Multiple Criteria Decision Making, edited by J. Cochrane and M. Zeleny, University of South Carolina Press, Columbia, South Carolina, 1973.

8. A. Papoulis, "Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, 1965, p. 183.

9. E. Parzen, "Modern Probability Theory and Its Applications, J. Wiley, New York, 1960, p. 318.

10. B. Roy, "How Outranking Relation Helps Multiple Criteria Decision Making," Selected Proceedings of a Seminar on Multiple Criteria Decision Making, edited by J. Cochrane and M. Zeleny, University of South Carolina Press, Columbia, South Carolina, 1973.

11. B. Roy, "Algèbre Moderne et Théorie des Graphes," Dunod, Paris, 1969.

12. B. Roy, "Problems and Methods with Multiple Objective Functions," Mathematical Programming I, 1971, pp. 239-266.

13. S. Watanabe, "Creative Learning and Propensity Automation," IEEE Trans. Syst., Man, Cybern., Vol. SMC-5, 1975, p. 306.

14. S. Watanabe, "Information-Theoretical Aspects of Inductive and Deductive Inference," IBM Journal of Research and Development, Vol. 4, April 1960, pp. 208-231.

15. S. Wanatabe, T. Fujii, A. Kamakura, "Experiments on Superinduction," IEEE, Systems, Man, and Cybernetics, Vol. SMC-8, No. 5, May 1978, pp. 401-42.

APPENDIX A

Unbiased Estimates of Probability Density Function

The entropy of information as suggested in section 5 is given by the following expression

$$E = -\int_{-\infty}^{\infty} p(f) \log_2 p(f)\, df \qquad \text{(A-1)}$$

where p(f) is the probability density function of f. f is a random variable characterizing the criterion under consideration. At this stage P(f) is unknown. It is determined according to the following maximization scheme (Maximum Entropy Criterion):

Maximize E subject to the following three constraints:

1. $$\int_{-\infty}^{\infty} p(f)\cdot df = 1 \qquad \text{(A-2)}$$

2. $$\int_{-\infty}^{\infty} p(f)\cdot f\cdot df = \overline{f} \qquad \text{(A-3)}$$

3. $$\int_{-\infty}^{\infty} p(f)\cdot(f-\overline{f})^2\cdot df = \sigma^2_f \qquad \text{(A-4)}$$

where $\overline{f}$ is the mean and σ_f^2 the variance of the random variable f.

The maximization of the entropy E given in equation (A-1) subject to three constraints is handled according to a variational constrained optimization scheme.

The Lagrangian equation is:

$$L = \int_{-\infty}^{\infty} p(f)\ \log_2 p(f)\ df + \lambda_1\left[\int_{-\infty}^{\infty} p(f)\ df - 1\right] + \lambda_2\left[\int_{-\infty}^{\infty} p(f)\cdot f\cdot df - \overline{f}\right] + \lambda_3\left[\int_{-\infty}^{\infty} p(f)\cdot(f-\overline{f})^2\ df - \sigma_f^2\right] \qquad \text{(A-5)}$$

where $p(f)$ is the unknown probability density function of f, $\overline{f}$ is the known mean value of f, σ_f^2 is the known variance of f and λ_1, λ_2, λ_3 are the Lagrange multipliers.

The maximum of L is obtained through Euler's procedure assuming that

$$p(f) = \hat{p}(f) + \epsilon_1\ \eta_1(f) \qquad \text{(A-6)}$$

where ϵ_1 are constants equal to zero for the maximum value of L and η_1 are arbitrary differentiable functions compatible with the constraints. Euler's equation is then the following:

$$-\frac{\partial}{\partial p}\left[p\ \log_2\ (p) + \lambda_1\ p + \lambda_2\ p\ f + \lambda_3\ p(f-\overline{f})^2\right] = 0 \qquad \text{(A-7)}$$

and finally it becomes

$$p(f) = e^{-\left[\log_2\ e\ +\ \lambda_1\ +\ \lambda_2\ f\ +\ \lambda_3(f-\overline{f})^2\right]} \qquad \text{(A-8)}$$

The normal distribution satisfies this expression and is substituted in the Entropy function

$$E = -\int_{-\infty}^{\infty} p(f)\ \log_2\ (p(f))\ df \qquad \text{(A-9)}$$

where $p(f) = \frac{1}{\sigma_f\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{f-\overline{f}}{\sigma_f}\right)^2}$