

AD-770 569

FLAT MAXIMA IN LINEAR OPTIMIZATION MODELS

Detlof V. Winterfeldt, et al

Michigan University

Prepared for:

Office of Naval Research  
Advanced Research Projects Agency

2 November 1973

DISTRIBUTED BY:

**NTIS**

National Technical Information Service  
U. S. DEPARTMENT OF COMMERCE  
5285 Port Royal Road, Springfield Va. 22151

AD770569

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER 011313-4-T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle)  FLAT MAXIMA IN LINEAR OPTIMIZATION MODELS		5. TYPE OF REPORT & PERIOD COVERED  Technical	
		6. PERFORMING ORG. REPORT NUMBER  None	
7. AUTHOR(s)  Detlof v. Winterfeldt and Ward Edwards		8. CONTRACT OR GRANT NUMBER (s)  N00014-67-A-0181-0049	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Engineering Psychology Laboratory Institute of Science & Technology University of Michigan Ann Arbor, Michigan 48105		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS  NR 197-021 ARPA Order No. 2105	
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Boulevard Washington, D. C. 22209		12. REPORT DATE  2 November 1973	
		13. NUMBER OF PAGES  31	
14. MONITORING AGENCY NAME AND ADDRESS (if different from Controlling Office) Engineering Psychology Programs Office of Naval Research Department of the Navy Arlington, Virginia		15. SECURITY CLASS (of this report)  Unclassified	
		15a DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release; distribution unlimited.			
17. DISTRIBUTION STATEMENT (of the information contained in this report, if different from Report)			
18. SUPPLEMENTARY NOTES  None			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  Linear optimization Flat maxima Utility Scoring rule Signal detection			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  Expected value functions as functions of decisions and decision strategies are flat around their maxima. This so called flat maximum phenomenon has been discovered in sensitivity analyses in virtually all decision theoretic paradigms. But until now most of the research on flat maxima explored more or less general examples and limiting considerations. Two basic questions remained unanswered: what are the mathematical reasons for the restricted shape of the evaluation functions; and can these restrictions be interpreted as flatness in a psychological sense? While the			

Reproduced by  
**NATIONAL TECHNICAL  
INFORMATION SERVICE**  
U S Department of Commerce  
Springfield VA 22151

second question calls for psychological experimentation, the first question can be answered with mathematical tools. The present article shows that the mathematical characteristics of linear optimization models impose severe restrictions on the functions evaluating choice alternatives such as gambles, multi-attributed outcomes, or consumption streams. The course of proof of this argument provides a helpful tool for sensitivity analyses in decision theory. The concepts and methods are demonstrated in examples from statistical decision theory, psychological modeling, and applied decision theory.

## Introduction

Decision theoretic models, such as expectation models, multi-attribute utility models, or time-discounting models, evaluate decisions or decision strategies with a real numbered index of preference such as a utility, a present value, risk, expected value, etc. Most prominent decision theoretic models are linear decomposition models that base their evaluation of choice alternatives (a generic term for gambles, multi-attributed outcomes, and consumption streams) on a weighted additive integration of subjective or objective input parameters, which the decision maker or experts provide by means of simple choices or judgmental tasks. For example, the subjective expected utility model evaluates gambles by combining subjective probabilities of events and utilities of decision outcomes into expected utilities.

One of the peculiarities of linear optimization models is the flatness of their evaluation function in the area of optimal choice alternatives. (We implicitly assume here and in the following that choice alternatives have a continuous or dense numerical description as vectors, decision functions, stopping rules, probability cutoffs, etc.) A suboptimal choice does not seriously hurt the decision maker as long as the alternative selected is not grossly away from the optimum. This type of insensitivity is closely linked to a second type, which is often found in decision analysis settings. Variations of model parameters like importance weights or subjective probabilities seldom produce drastic changes in the model evaluation function. A set of quite different parameter values may lead to the selection of the same choice alternative; and even if the use of a wrong set of parameter values leads to a different decision, the

first type of insensitivity will guarantee that the loss in expected value as calculated by means of the model with the correct parameters will be rather small. Some researchers (Yntema and Torgerson, 1961) have even argued for an insensitivity across models. According to their results different models should—under some mild conditions—lead to similar evaluations and decisions.

Although there are doubts about insensitivity across models (see Fischer, 1972) the evidence for the two other kinds of insensitivity is substantial. In expectation models v. Winterfeldt and Edwards (1973) generalized scattered findings of flat expected value functions as functions of decisions and decision strategies. In multi-attribute utility theory Fischer (1972) demonstrated the insensitivity of multi-attribute utility functions against variations in parameters like importance weights and single dimension utilities.

But up until now the evidence for flat maxima was based on more or less general examples. The questions remained whether or not flatness is a necessity and what model characteristics cause it. Another problem with the arguments for insensitivity in those examples is the concept of flatness itself. A function may look flat, but that can easily be fixed by stretching the units of the ordinate and compressing the units of the abscissa. Flatness is not a mathematical, but a psychological concept. 5% loss may be substantial for one decision maker and negligible for another.

These arguments call for two kinds of research on the flat maximum phenomenon: first, a mathematical analysis that proves the inevitability of restricted forms of the evaluation functions, given certain model characteristics, and second, an experimental psychological analysis that shows whether or not these restrictions can be interpreted as flatness.

This report presents the mathematical foundation of the flat maximum phenomenon. Integrating some theorems from statistical decision theory it shows that the nature of all linear optimization models imposes severe restrictions on the model evaluation function. The mathematical proofs produce two further important and practical results: they establish an equivalence between model insensitivity against variations in choice alternatives and against variations in parameter values; and they present the tools for a general and simple approach to sensitivity analyses. Some examples from statistics, psychological modeling, and decision analysis demonstrate the use of the concepts and methods developed.

#### Why Evaluation Functions Are Restricted

The most severe restriction on a function is, of course, the specification of its functional form and parameters, which determines each point of its graph. At the other extreme one may only know that  $f$  is a function. Between these extremes there are more or less severe confining properties such as convexity, continuity, boundedness, number of minima and maxima, etc. Assume, for example, that all we know about the function  $y = f(x)$  is

- (a) it is defined for  $0 \leq x \leq 1$  and bounded between  $y = 0$  and  $y = 1$ ;
- (b) it is strictly convex;
- (c) it is continuous;
- (d) it has a unique minimum.

Figure 1a gives some examples of graphs of functions which satisfy (a) - (d).

Figure 1b illustrates some inadmissible cases.

-----  
Insert Figures 1a and 1b about here  
-----

This section will present the mathematical proof that evaluation functions in linear optimization models have confining properties like the ones discussed in this example. The substance of our argument are three theorems from statistical decision theory, which are proven in Ferguson (1967) and DeGroot (1970). The arguments and proofs are quite technical, but all theorems have a simple intuitive meaning, and except for theorem 3 they seem self evident. Rather than boring the reader with messy mathematics, we will rely on self-evidence, whenever possible, and confine ourselves to interpretation. The reader interested in more mathematical detail should consult the two references cited. For illustration a scoring rule example will accompany all theorems and proof arguments.

We want to study the behavior of the model evaluation function  $U$  which is defined over a set of choice alternatives  $X = [\underline{x}, \underline{y}, \underline{z}, \dots]$ . For example,  $X$  may be a set of gambles, decision functions, or multi-attributed outcomes;  $U$  may be a utility function or an expected utility function. In our scoring rule example  $X$  will be a set of probability estimates which are gambles by the definition of a scoring rule;  $U$  is the subjective expected value (SEV) of those gambles.

The application of a linear decomposition model to such a choice situation requires each  $\underline{x}$  to be described as an  $n$ -tuple of elements  $x_i$ , characterizing  $\underline{x}$  for a specific aspect or state  $S_i$  of the choice situation. We assume therefore that  $\underline{x}$  has the following representation:

$$\underline{x} = (x_1, x_2, \dots, x_i, \dots, x_n)$$

For example,  $\underline{x}$  may be a gamble in which one receives a dollar amount  $x_i$  if event  $S_i$  occurs; or a multi-attribute outcomes with value  $x_i$  in attribute  $S_i$ , or a cash flow in which one receives a dollar amount  $x_i$  at time  $S_i$ . Note that by labelling we implicitly let the number of states be finite. This finiteness of the state space will be our first assumption (A1) for the further mathematical development.

Linear decomposition models go further by defining utility functions  $u_i$  within each state  $S_i$  so that each choice alternative can now be characterized by a vector of single state utilities:

$$\underline{u}(\underline{x}) = (u_1(x_1), u_2(x_2), \dots, u_i(x_i), \dots, u_n(x_n))$$

According to our second assumption (A2) these utilities are bounded, i.e.,  $m \leq u_i(x_i) \leq M$  for all  $i$  and  $\underline{x} \in X$ , and some real  $m, M$ .

Furthermore, in linear optimization models a weight vector  $\underline{w}$  from a parameter set  $W$  associates with each state  $S_i$  a weight  $w_i$  which can be interpreted as a subjective probability, an importance weight, or a discounting rate. In expectation, time discounting, and multi-attribute models we can assume that  $w_i \geq 0$  and  $\sum_{i=1}^n w_i = 1$ .

The linear model evaluates choice alternatives  $\underline{x}$  now by computing the scalar product  $\underline{w} \circ \underline{u}$  of the vectors  $\underline{u}$  and  $\underline{w}$ , or more simply as the weighted average:

$$U(\underline{x}, \underline{w}) = \underline{w} \underline{ou}(\underline{x}) = \sum_{i=1}^n w_i u_i(x_i) \quad (1)$$

That alternative  $\underline{x}^*$  is optimal, which maximizes<sup>1</sup>  $U$ , i.e., the decision rule of linear optimization models is:

"choose  $\underline{x}^*$  with  $U(\underline{x}^*, \underline{w}) \geq U(\underline{x}, \underline{w})$  for all  $\underline{x} \in X$ "

We will define

$$U^*(\underline{w}) = U(\underline{x}^*, \underline{w}) \quad (2)$$

i.e.,  $U^*$  is the maximal attainable utility for a specific weight vector  $\underline{w}$ . In statistical decision theory  $\underline{x}^*$  would be called a Bayes decision with respect to the prior distribution  $\underline{w}$ .

Let us interpret the previous paragraph in the scoring rule situation, assuming a simple two state case, in which  $S_1$  and  $S_2$  are two mutually exclusive and exhaustive events,  $w_1$  and  $w_2$  are the associated true subjective probabilities (SP's). The set of choice alternatives  $X$  is a subset of the real plane  $R^2$ , namely the tuples  $(x_1, x_2)$  with  $0 \leq x_1 \leq 1$  and  $x_1 + x_2 = 1$ . The  $x_i$ 's are interpreted as the stated probabilities of the events  $S_i$ . Since  $x_1 = 1 - x_2$  the choice set can be totally characterized by the real numbers between 0 and 1. Scoring functions  $u_1$  and  $u_2$  are defined for each state  $S_i$  such that  $U(\underline{x}, \underline{w}) = SEV(\underline{x}, \underline{w})$  is maximal if  $\underline{x} = \underline{w}$ . We will specifically analyze the quadratic scoring rule in which:

<sup>1</sup>Our whole argument will be based on maximization. The dual argument based on minimization is basically the same.

$$u_1(x_1) = 1 - (1 - x_1)^2 \quad (3a)$$

$$u_2(x_2) = 1 - x_1^2 \quad (3b)$$

Schematically the scoring rule paradigm is represented in Table 1.

-----  
 Insert Table 1 about here  
 -----

Here, of course,  $U^*(\underline{w})$  has a very clear interpretation:  $U^*(\underline{w}) = U(\underline{x}^*, \underline{w}) = U(\underline{w}, \underline{w})$ .

Before we enter into a discussion of the behavior of  $U$  and  $U^*$ , we need to state two preliminary theorems, which will establish a relation between the parameter set  $W$  and the choice set  $X$ .

THM 1 Assuming that the state space is finite (A1) and that the

$u_i(x_i)$  are bounded (A2), there exists for every  $\underline{w} \in W$  at

least one  $\underline{x} \in X$  such that  $U(\underline{x}, \underline{w}) = U^*(\underline{w})$ .

We will define the subset of  $X$ , which contains those elements  $\underline{x}$  which are optimal under  $\underline{w}$  as  $X_{\underline{w}}$ , and elements of  $X_{\underline{w}}$  as  $\underline{x}_{\underline{w}}$ . A similar theorem is proven in Ferguson (1967). It seems self evident for finite  $X$ : you just order the  $\underline{x}$ 's according to their  $U$ -values (all of which are finite by A1 and A2) and choose the  $\underline{x}$  with the maximal  $U$ .

The second theorem is more sophisticated and, in fact, substantial work is based on it in decision theory. To state it, we first have to introduce the notion of dominance (here in a somewhat wider sense than usual). We call a choice alternative  $\underline{x}$  dominated, if there exist other alternatives  $\underline{y}$  and  $\underline{z}$  and a real number  $a$  ( $0 \leq a \leq 1$ ) such that

$$(1) \quad u_i(x_i) \leq au_i(y_i) + (1-a)u_i(z_i) \quad \text{for all } i \quad (4a)$$

and

$$(2) \quad u_i(x_i) < au_i(y_i) + (1-a)u_i(z_i) \quad \text{for some } i. \quad (4b)$$

The set of non-dominated alternatives is called admissible. We label the admissible subset of  $X$  as  $\tilde{X}$ , and we will assume in the following that  $X = \tilde{X}$  (A3).

THM 2 Given A1, A2, and A3, there exists for all  $\underline{x} \in X$  at least one  $\underline{w} \in W$  such that  $U(\underline{x}, \underline{w}) = U^*(\underline{w})$ .

Similarly to theorem 1 we will define the subset of  $W$  which contains those parameters  $\underline{w}$  which would make  $\underline{x}$  an optimal choice  $W_{\underline{x}}$ , and elements  $\underline{w} \in W_{\underline{x}}$  we will call  $\underline{w}_{\underline{x}}$ . This theorem is rather difficult to prove and requires a substantial number of "lemmas" such as the famous separating hyperplane theorem. The idea of theorem 2, however, is simple: admissible choice alternatives are potential candidates for optimal choices.

Theorems 1 and 2 allow us to step freely from the parameter set  $W$  to the choice set  $X$  and back in our analysis of  $U$  as a function of both,  $\underline{w}$  and  $\underline{x}$ . The main purpose of these theorems here is to establish an equivalence between parameters and choice alternatives for the insensitivity analysis.

Both theorems have a simple interpretation in our scoring rule example. Since here  $X = W$  and, by definition of a proper scoring rule  $X = \tilde{X}$ , the theorems say that for each true subjective probability vector  $\underline{w}$ , there is an optimal probability estimate  $\underline{x}$ , and for each estimate  $\underline{x}$  there is a subjective probability vector  $\underline{w}$  which would make this estimate optimal. In fact, we already knew that, since the unique value  $\underline{x} = \underline{w}$  was the best estimate in the SEV sense.

After these preliminary theorems we are now able to study the restrictions on  $U$  and  $U^*$ . By the definition of  $U^*$  and by the properties of linear optimization models, we know that

- (1) the range of  $U^*$  will be the restricted range of  $W$ ,
- (2)  $U^*$  has to go through all the corner points  $[\sup_X \{u_i(x_i)\}, w_i = 1]$ .

But our third theorem imposes a much more severe restriction on  $U^*$ :

THM 3 Under A1 and A2  $U^*$  as defined in (3) is a convex function of  $\underline{w}$ , i.e.,  $U^*[a\underline{v} + (1-a)\underline{w}] \leq aU^*(\underline{v}) + (1-a)U^*(\underline{w})$  for all  $0 \leq a \leq 1, \underline{v}, \underline{w} \in W$ .

The proof is rather simple, and it is presented here, since the convexity of  $U^*$  is not at all self evident. For a different version of the proof, see DeGroot (1970). Consider the vector  $a\underline{v} + (1-a)\underline{w}$ . From theorem 1 we know that there is at least one  $\underline{x}$  such that

$$U[\underline{x}, a\underline{v} + (1-a)\underline{w}] = U^*[a\underline{v} + (1-a)\underline{w}]. \quad (5)$$

By definition of  $U$

$$U[\underline{x}, a\underline{v} + (1-a)\underline{w}] = [a\underline{v} + (1-a)\underline{w}] \circ u(\underline{x}) = a\underline{v} \circ u(\underline{x}) + (1-a) \underline{w} \circ u(\underline{x}). \quad (6)$$

The latter equality follows from the distributivity of " $\circ$ ." Again by theorem 1 there exist  $\underline{y}$  and  $\underline{z} \in X$  such that

$$U(\underline{y}, \underline{v}) = U^*(\underline{v}) \quad (7)$$

and

$$U(\underline{z}, \underline{w}) = U^*(\underline{w}). \quad (8)$$

Since

$$U^*(\underline{v}) \geq U(\underline{x}, \underline{v}) = \underline{v} \underline{u}(\underline{x}) \quad (9)$$

and

$$U^*(\underline{w}) \geq U(\underline{x}, \underline{w}) = \underline{w} \underline{u}(\underline{x}) \quad (10)$$

it follows by substitution that

$$U^*[a\underline{v} + (1-a)\underline{w}] = aU(\underline{x}, \underline{v}) + (1-a)U(\underline{x}, \underline{w}) \leq aU^*(\underline{v}) + (1-a)U^*(\underline{w}). \quad (11)$$

What does this theorem mean in our scoring rule example? Defining  $U^*(\underline{w})$  as  $U^*(w_1)$ , we see that  $U^*$  is severely restricted through the boundaries and by convexity. Figure 2a gives some examples of graphs of  $U^*$  functions which might have been generated by some scoring rule (actually,  $U^*$  is equivalent to some

-----  
 Insert Figures 2a, 2b, and 2c about here  
 -----

scoring rule). Figure 2b shows inadmissible graphs. Figure 2c shows the  $U^*$  function for our quadratic scoring rule. The interpretation of convexity in this example is very intuitive: the more certain you are about the events  $S_1$ , the better your optimal decision will be in terms of SEV.

We know now that  $U^*$  is a restricted function of  $\underline{w}$ , but what about  $U$  as a function of  $\underline{x}$ ? With theorems 1 and 2 it becomes simple to step from  $U^*$  to  $U$ .  $U$  has two arguments,  $\underline{w}$  and  $\underline{x}$ . We know that  $U$  is linear in the  $w_1$ 's, thus as a function of  $\underline{w}$   $U$  defines an  $n-1$  dimensional hyperplane. In the scoring rule

case  $U$  is a line as illustrated in Figure 3.

-----  
Insert Figure 3 about here  
-----

What do these lines, planes, and hyperplanes have to do with  $U^*$ ? First,  $U$  is defined on the same space  $W$  on which  $U^*$  is defined. Second,  $U(\underline{x}, \underline{w})$  and  $u^*(\underline{w})$  have at least one point in common, namely the point  $[U^*(\underline{w}), \underline{w}]$ . Third,  $U^*$  is everywhere at least as large as  $U$ , i.e.,  $U^*(\underline{w}) \geq U(\underline{x}, \underline{w})$  for all  $\underline{w} \in W$  and  $\underline{x} \in X$ . This last fact follows by simple contradiction. If  $U^*$  was not at least as large as  $U$  for all  $\underline{w}$  and  $\underline{x}$ , then there would exist some  $\underline{x}$  and  $\underline{w}$  such that  $U(\underline{x}, \underline{w}) > U^*(\underline{w})$ , which contradicts the definition of  $U^*(\underline{w})$ . Therefore, as a line  $U$  is a tangent to  $U^*$ , as a plane it is a tangent plane, and as a hyperplane it is a tangent hyperplane to  $U^*$ . Figure 4 clarifies these concepts in our scoring rule example.

Figure 4 also exemplifies how the restrictions on  $U$  and the possible losses  $\Delta U$  are determined totally by the shape and the slopes of  $U^*$ . All losses which

-----  
Insert Figure 4 about here  
-----

may be encountered in a choice situation (whether they are due to a suboptimal choice or the use of a wrong set of parameter values) are differences between  $U^*$  and some hyperplane tangent to it. Assume in a two state case you could construct  $U^*$  without restrictions and you wanted to make losses around a value  $\underline{z}$  as large as possible within the boundaries of  $u_1$ . You probably would construct a  $U^*$  function which looks somewhat like in Figure 5. But by convexity of

-----  
Insert Figure 5 about here  
-----

$U^*$  this shape is inadmissible. The convexity of  $U^*$  will make losses in the area of the optimal choice alternative small.

This intuitive interpretation of the restrictions on the behavior of  $U$  around its maximum through the convexity of  $U^*$  can also be expressed mathematically. Since  $U$  is a tangent hyperplane to  $U^*$ , it can be totally determined by  $n-1$  slopes and one point. Assuming that  $U^*$  is differentiable at  $\underline{w}_x$ , the actual formula for  $U$  in terms of  $U^*$  is consequently:

$$U(\underline{x}, \underline{w}) = U^*(\underline{w}_x) + \sum_{i=1}^{n-1} d_i (w_i - w_{ix}) \quad (12)$$

where  $d_i = du^*/dw_i |_{\underline{w}_x}$ , i.e., the directional slope of  $U^*$  evaluated at  $\underline{w}_x$ .

How much do we stand to lose by the choice of a nonoptimal alternative? Assume that  $\underline{w}_y$  is the true weight vector,  $\underline{y}$  is the optimal choice alternative, but instead we choose  $\underline{x} \neq \underline{y}$ . Since by definition

$$U(\underline{y}, \underline{w}_y) = U^*(\underline{w}_y) \quad (13)$$

and

$$U(\underline{x}, \underline{w}_y) = U^*(\underline{w}_x) + \sum_{i=1}^{n-1} d_i (w_{iy} - w_{ix}). \quad (14)$$

we will lose

$$\Delta U = U^*(\underline{w}_y) - U^*(\underline{w}_x) - \sum_{i=1}^{n-1} d_i (w_{iy} - w_{ix}). \quad (15)$$

The convexity of  $U^*$  puts limits on the differences between the  $U^*$ 's as well as on the slopes  $d_i$ . Since, in addition  $(w_{iy} - w_{ix})$  cannot exceed 1 (and will typically be much smaller) the loss  $\Delta U$  will remain small.

How much will we lose if we base our decision on a parameter value  $\underline{y}$  when, in fact, the true value is  $\underline{w}$ ? We would choose  $\underline{x}$  such that

$$U(\underline{x}, \underline{y}) = U^*(\underline{y}) \quad (16)$$

We will receive

$$U(\underline{x}, \underline{w}) = U^*(\underline{y}) + \sum_{i=1}^{n-1} d_i (w_i - v_i) \quad (17)$$

and consequently we will lose

$$\Delta U = U^*(\underline{w}) - U^*(\underline{y}) - \sum_{i=1}^{n-1} d_i (w_i - v_i). \quad (18)$$

Two general expressions may be helpful for limiting purposes: the maximum possible loss is determined by

$$\Delta U_{\max} = \max_{k, l} \{U^*(\underline{e}_k) - U^*(\underline{f}_l) - d_k + d_l\} \quad (19)$$

where  $\underline{e}_k$  and  $\underline{f}_l$  are the unit weight vectors with  $e_k = 1$  and  $f_l = 1$ . See Figure 6 for illustration.

-----  
 Insert Figure 6 about here  
 -----

But this loss would only result from an extremely foolish choice. By choosing  $\underline{y}$  such that the value  $U(\underline{y}, \underline{w})$  is the minimal point of  $U^*$  (in decision theoretic terms  $\underline{y}$  is the minimax strategy), we can reduce the maximum possible loss to

$$\Delta U_{\minimax} = \sup_X \sup_i \{u_i(x_i)\} - U^*(\underline{w}). \quad (20)$$

See Figure 7 for illustration.

-----  
Insert Figure 7 about here  
-----

By now it should be clear how to do a sensitivity analysis with the tools developed. The first step is to construct the function  $U^*$ . Often this can be done explicitly. If an explicit solution is not possible or too difficult and time consuming, one can help oneself with the following procedure: first, plot the cornerpoints  $U^*(\underline{e}_k)$ , then determine  $U^*(\underline{w})$  where  $\underline{w}$  is the "least favorable" weight vector which would make a minimax choice optimal. Then find some other points of  $U^*$  and exploit the convexity property to approximate the whole function. Alternatively  $U^*$  can be approximated by plotting some  $U$  - lines. This procedure can be done graphically in two state cases. In cases with a larger number of states computer aid is needed. Equations (19) and (20) give some boundary losses, and equation (15) determine for each particular case the potential losses. In general: the flatter  $U^*$  as a function of  $\underline{w}$ , the flatter  $U$  as a function of  $\underline{x}$  will be around its maximum.

To summarize this section: First we established a relation between the parameter set and the choice set in two theorems by making three assumptions. We assumed that the state space is finite (A1), that the single state utility functions are bounded (A2), and that the choice set is admissible (A3). Then we showed that under A1 and A2 in linear optimization models the function  $U^*$  is severely restricted by its boundaries and through convexity. Finally, we demonstrated the restrictions on the actual evaluation function  $U$  as a function of  $U^*$  and outlined a general approach to sensitivity analysis using the properties of  $U^*$ .

The next section will give some examples to demonstrate the concepts and methods developed.

### Examples

#### A Signal Detection Example

We assume a simple two state signal detection situation in which a datum  $d$  may be sampled from either of two normally distributed populations  $S_1$  or  $S_2$ . These distributions have equal variances  $s = 1$  and different means  $m_1$  and  $m_2$ . Two decisions can be made upon observing  $d$ :

(1)  $a_1$ :  $d$  was sampled from  $S_1$ ,

or

(2)  $a_2$ :  $d$  was sampled from  $S_2$ .

The prior probability for sampling from  $S_1$  is  $w_1$ , with  $w_1 = 1 - w_2$ . Payoffs are 1 for correct decisions, 0 for incorrect decisions.

The choice set  $X$  here is the set of real valued decision functions  $x$ , which are cutoffs along the possible real values of  $d$  ( $x$  is in this case related to the usual likelihood ratio criterion  $\beta$  by  $x = \ln\beta/d'$ ).  $x$  is evaluated by a simple expected value model.

To formulate the problem in the format of the preceding section, we construct the expected value matrix, where expectations are taken over the random variable  $d$  within each state  $S_i$ . This matrix indicates for each  $x$  the expected

-----  
Insert Table 2 about here  
-----

amount of money the decision maker stands to lose under  $S_1$  (see Table 2). The expected values are defined as

$$EV(x|S_1) = \Pr(d < x|S_1) \quad (21)$$

$$EV(x|S_2) = \Pr(d > x|S_2). \quad (22)$$

As in the scoring rule example, we have in this paradigm a 1:1 mapping from prior probabilities  $w_1$  into the choice alternatives  $x$ . As it is well known  $\beta^*$ , the optimal likelihood ratio criterion for the payoffs given is

$$\beta^* = w_1/(1-w_1) \quad (23)$$

and consequently

$$x^* = \ln [w_1/(1-w_1)]/d' \quad (24)$$

where  $x^*$  is the optimal cutoff value under  $w_1$ , or  $w_1 \in W_{x^*}$ .

-----  
Insert Figure 8 about here  
-----

Figure 8 shows the  $U^*$  function as determined from Table 2.

$$U^*(w_1) = w_1 \Pr(d < x^*/S_1) + (1-w_1) \Pr(d > x^*/S_2). \quad (25)$$

On the abscissa we have ordered the  $x^*$ -values under  $w_1$  to show how they are related.

Assume now that  $w_1 = .5$  is the true prior probability, but instead of

$x^* = 0$  we choose some other  $x' = +\infty$  which would be optimal under  $w_1 = 1$ .

Figure 8 demonstrates the possible loss  $\Delta U$  we expect in this case.

We see how the flatness of  $U^*$  prevents this loss from being large. v. Winterfeldt and Edwards (1972) showed in a direct analysis of the U-function, that U is generally flat in signal detection situations.

### A Multi-Attribute Example

Assume that we have two attributes on which we evaluate riskless options, say job offers. Attribute  $S_1$  may be salary,  $S_2$  may be staff benefits. We have five offers, each of which has been evaluated by a utility function  $u_i$  in each

-----  
Insert Table 3 about here  
-----

attribute (see Table 3). We can immediately delete  $x_3$  since it is dominated:

$$\underline{u}(\underline{z}) = \frac{1}{2} \underline{u}(x_2) + \frac{1}{2} \underline{u}(x_4) = (8, 10) \quad (26)$$

i.e.,

$$u_1(z_1) > u_1(x_{31}) \quad \text{and} \quad u_2(z_2) \geq u_2(x_{32}) \quad (27)$$

All other alternatives are admissible.  $U^*$  in this case will be piecewise linear, and its construction is rather easy. We just plot all the functions

$$U(\underline{x}_i, w_1) = w_1 u_1(x_{i1}) + (1-w_1) u_2(x_{i2}) \quad (28)$$

(see Figure 9). Naturally  $U^*$  is defined by the line segments of U such that

$$U(\underline{x}_i, w_1) \geq U(\underline{x}_j, w_1) \quad \text{for all } j. \quad (29)$$

(in Figure 9 marked by the solid line). Assume now that we choose  $\underline{x}_5$  for

-----  
Insert Figure 9 about here  
-----

$w_1 = 1/2$ . Figure 9 also indicates what we will stand to lose.

Similar analyses can be done with any matrix like the one in Table 3, as we find them in time discounting models or simple decision analysis problems. For more than two states graphical representations become impossible, and computer aid is needed. In those cases one should use the approach of bounding losses by the slopes and points of  $U^*$  as sketched in the previous section.

## References

DeGroot, M.H. Optimal Statistical Decisions. New York: McGraw-Hill, 1970.

Ferguson, T.S. Mathematical Statistics, A Decision Theoretic Approach. New York, Academic Press, 1967.

Fischer, G. W. Four Methods for Assessing Multi-attribute Utilities. Technical Report No. 03/230-6-T, Engineering Psychology Laboratory, The University of Michigan, 1972.

Yntema, D. B. and Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. IRE Transactions on Human Factors in Electronics, 1961, HFE-2, 20-26.

v. Winterfeldt, D. and Edwards, W. Costs and Payoffs in Perceptual Research. Technical Report No. 011313-1-T, Engineering Psychology Laboratory, The University of Michigan, 1973.

Table 1

Schematical representation of the scoring rule situation with quadratic scoring functions ( $SEV(x_1) = w_1[1-(1-x_1)^2] + (1-w_1)(1-x_1^2)$ ).

Probability Estimates  $x_1 = Pr(S_1)$

True SP's	$w_1$	$w_2$
States	$S_1$	$S_2$
0.0	0	1
⋮	⋮	⋮
$x_1$	$1-(1-x_1)^2$	$1-x_1^2$
⋮	⋮	⋮
0.5	.75	.75
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
1.0	1	0

Table 2

Expected value table for the two state signal detection situation ( $m_1 = -.5$ ;  $m_2 = +.5$ )

Decision Cutoff  $x = \ln \beta/d'$

Prior Prob.	$w_1$	$w_2$
States	$S_1$	$S_2$
$-\infty$	0	1
⋮	⋮	⋮
$x$	$\Pr[d < x   S_1]$	$\Pr[d > x   S_2]$
⋮	⋮	⋮
0	.69	.69
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
$+\infty$	1	0

Table 3

Multi-attributed outcomes  $x_i$  described  
by their single attribute utilities

Importance Weight	$w_1$	$w_2$
	$s_1$	$s_2$
$x_1$	2	12
$x_2$	6	11
$x_3$	7	10
$x_4$	10	9
$x_5$	12	5

Outcomes

Figure 1a

Graphs of functions which satisfy (a) - (d)

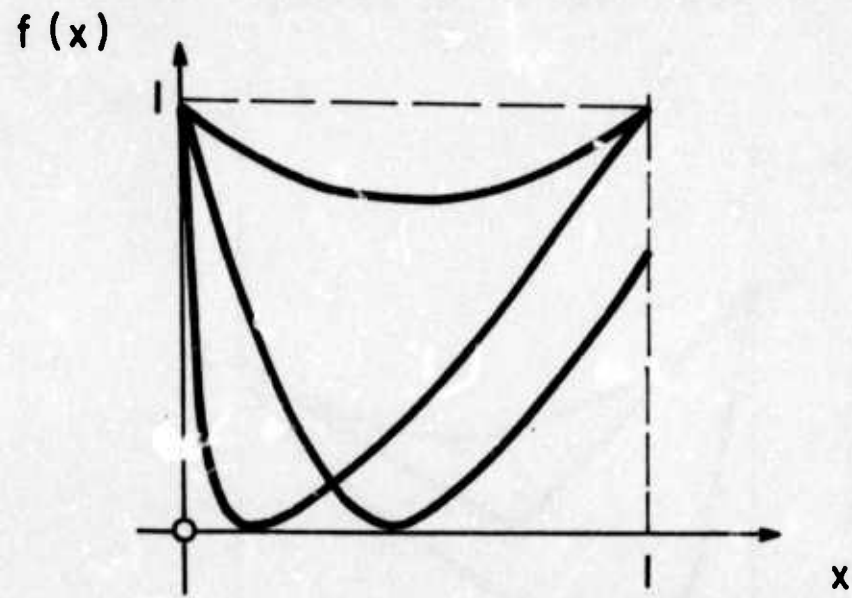


Figure 1b

Graphs of functions which do not satisfy (a) - (d)

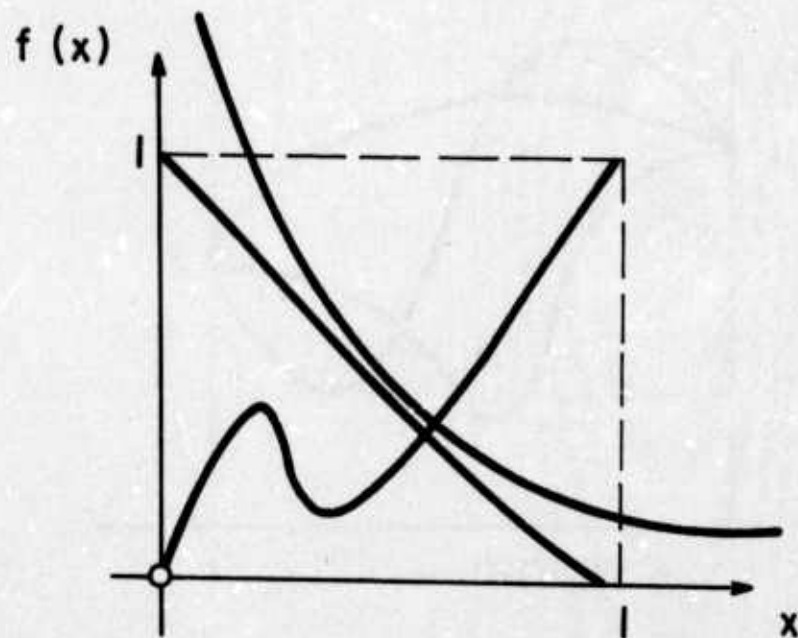


Figure 2a

Graphs of hypothetical  $U^*$ -functions which satisfy the boundary conditions  $U^*(0)=k$  and  $U^*(1) = 1$  and convexity.

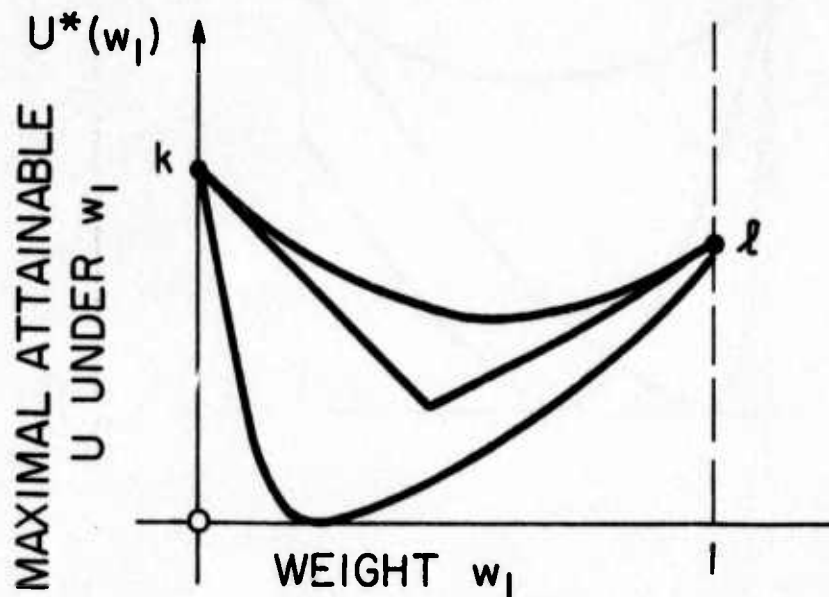


Figure 2b

Graphs of non-admissible  $U^*$  functions

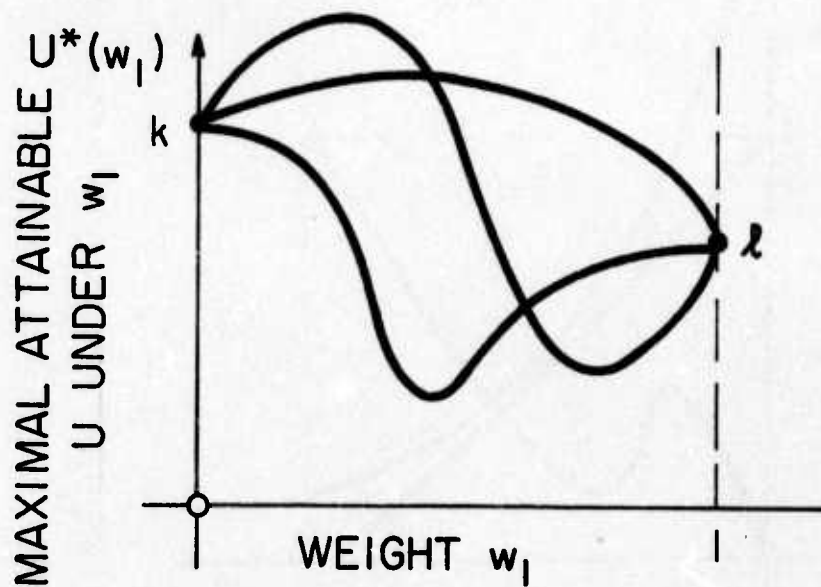


Figure 2c

U\*-functions in our scoring rule example

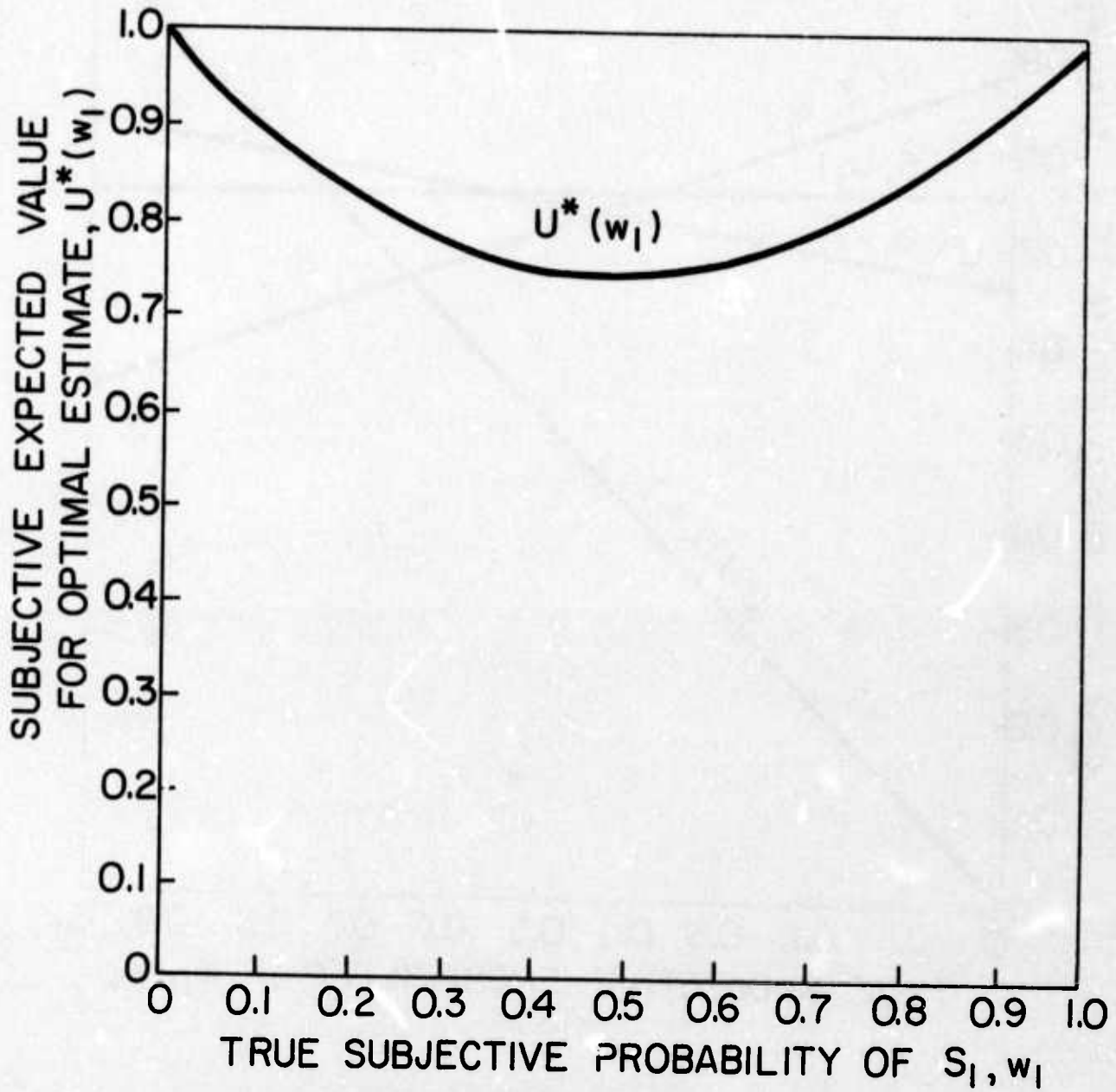


Figure 3

The lines defined by  $U(\underline{x}, \underline{w})$  in the scoring rule example

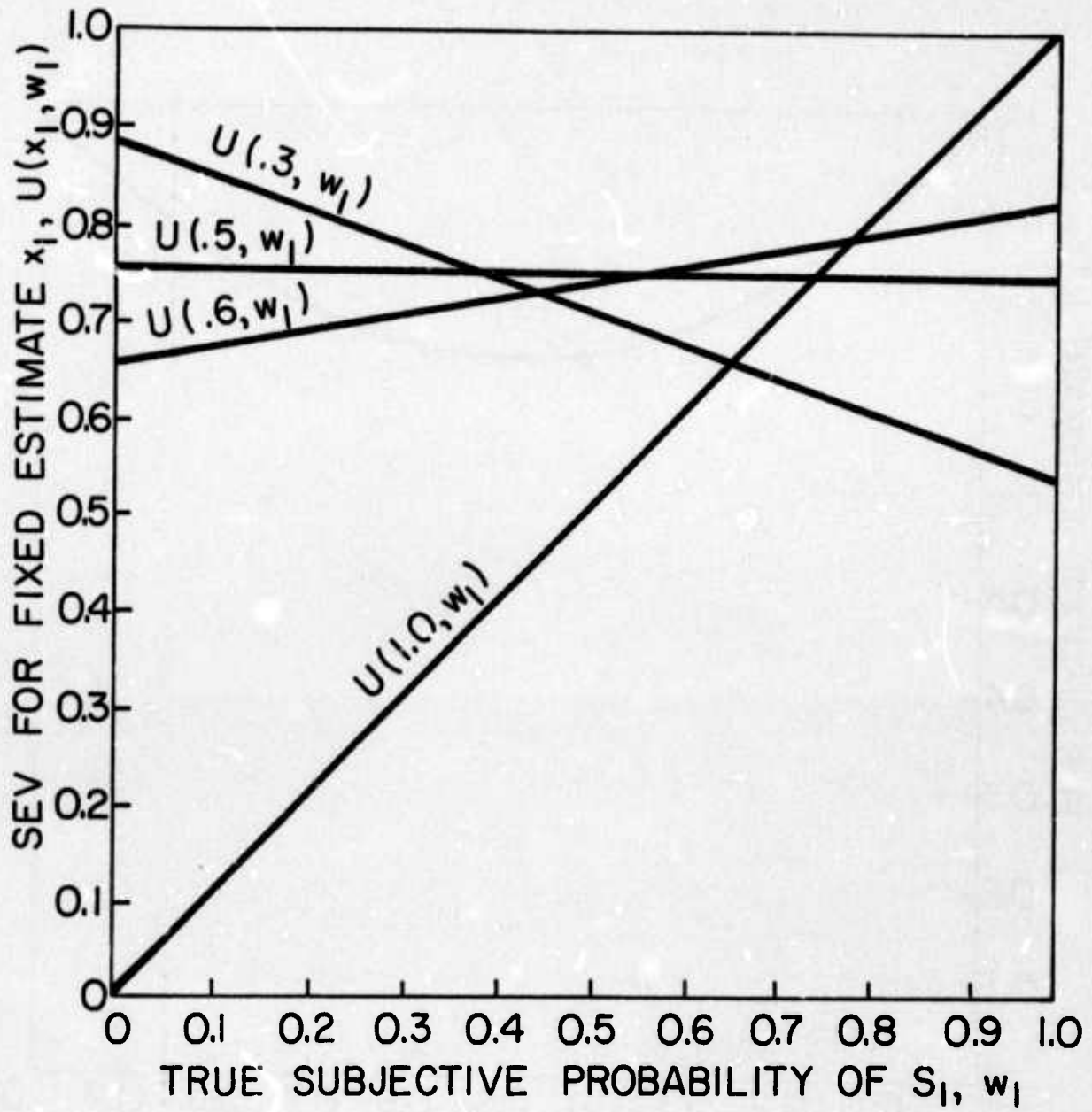


Figure 4

Demonstration of the relation between the hyperplanes  $U(\underline{x}, \underline{w})$  and  $U^*(\underline{w})$  in the scoring rule example

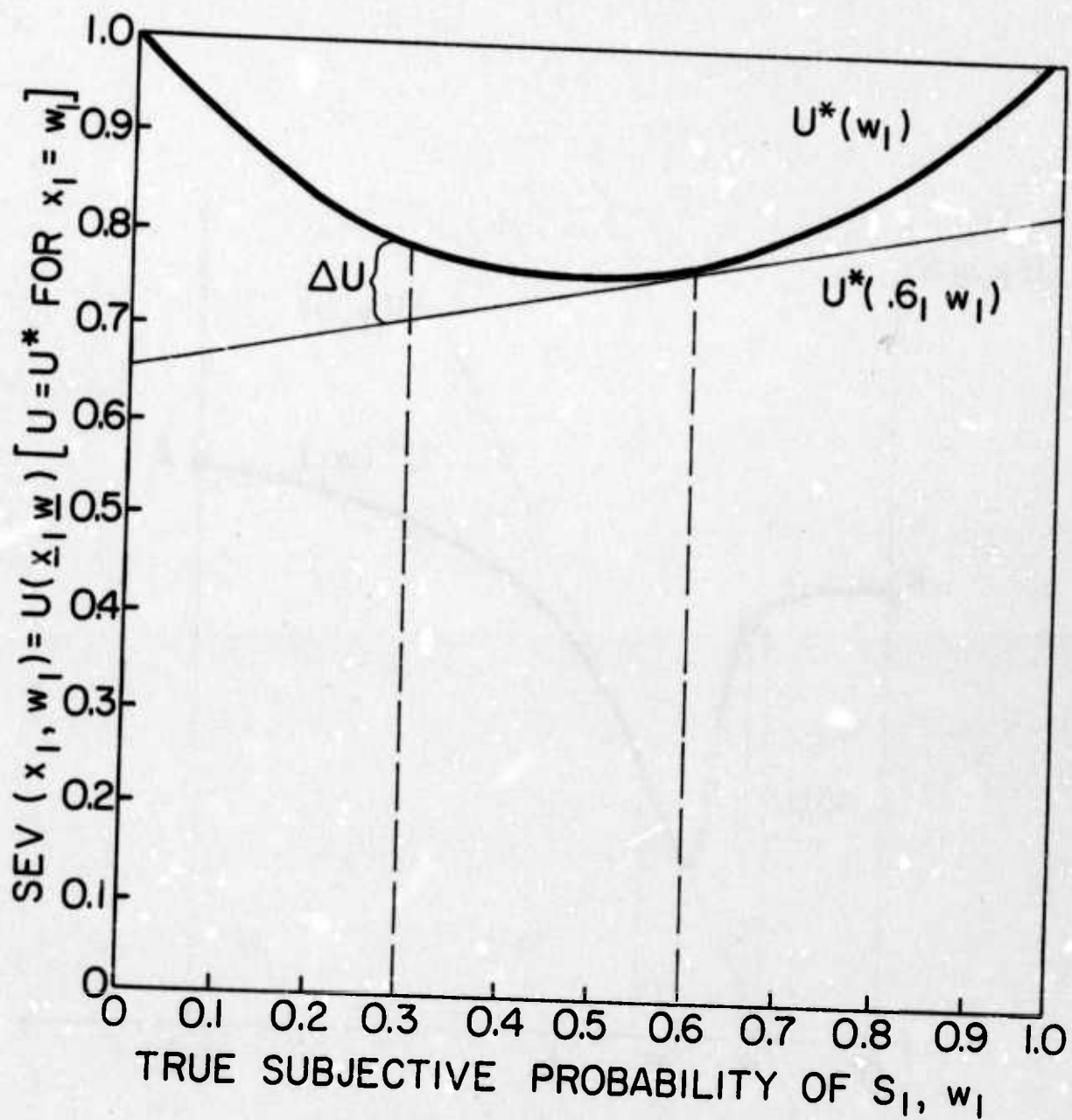


Figure 5

Inadmissible  $U^*$  function (with boundary values  $k$  and  $l$ ) which would make the loss due to small deviations from an optimal choice severe.

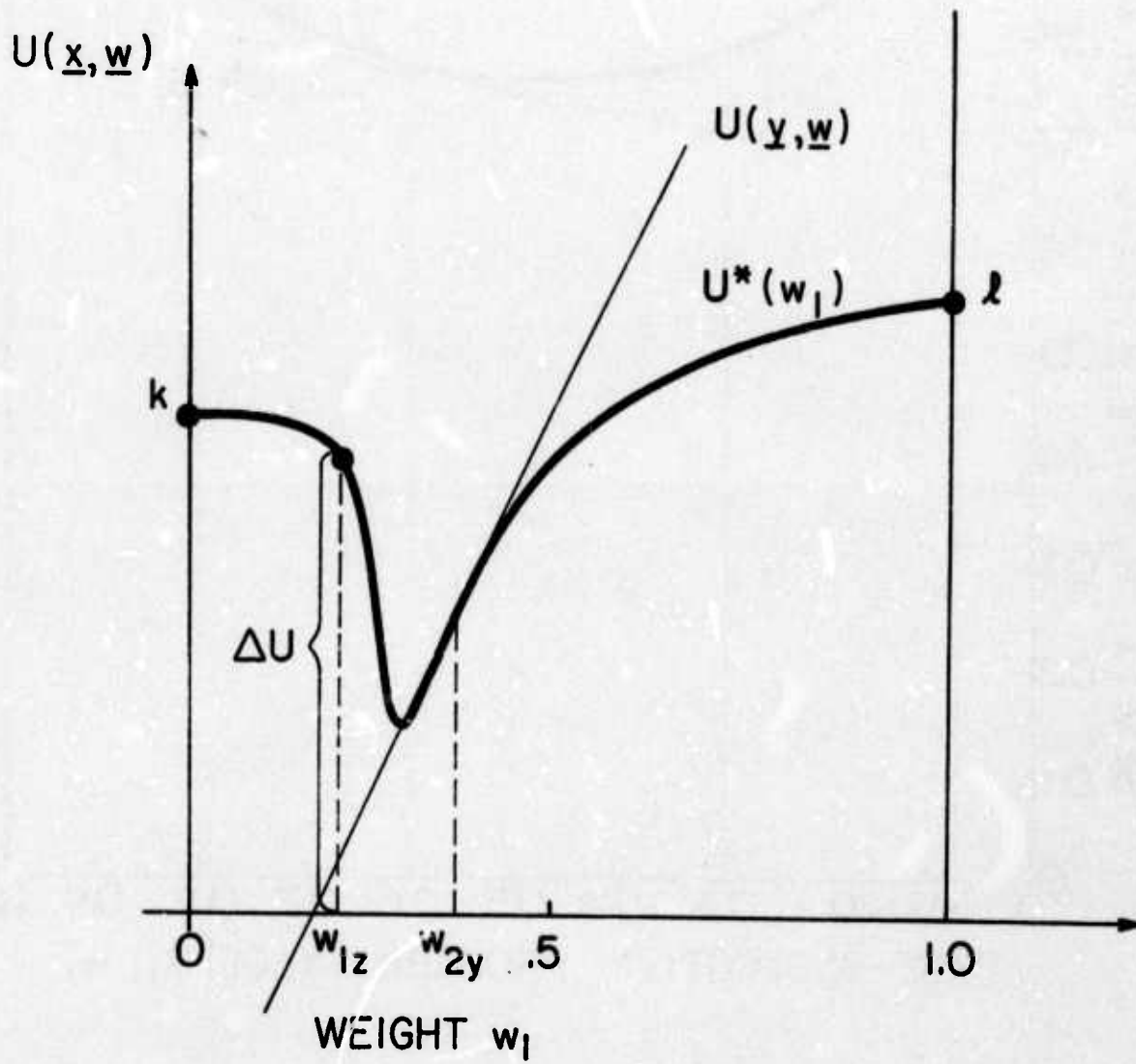


Figure 6

Graphical determination of maximum possible loss  
 $\Delta U_{\max}$  in a two state case.

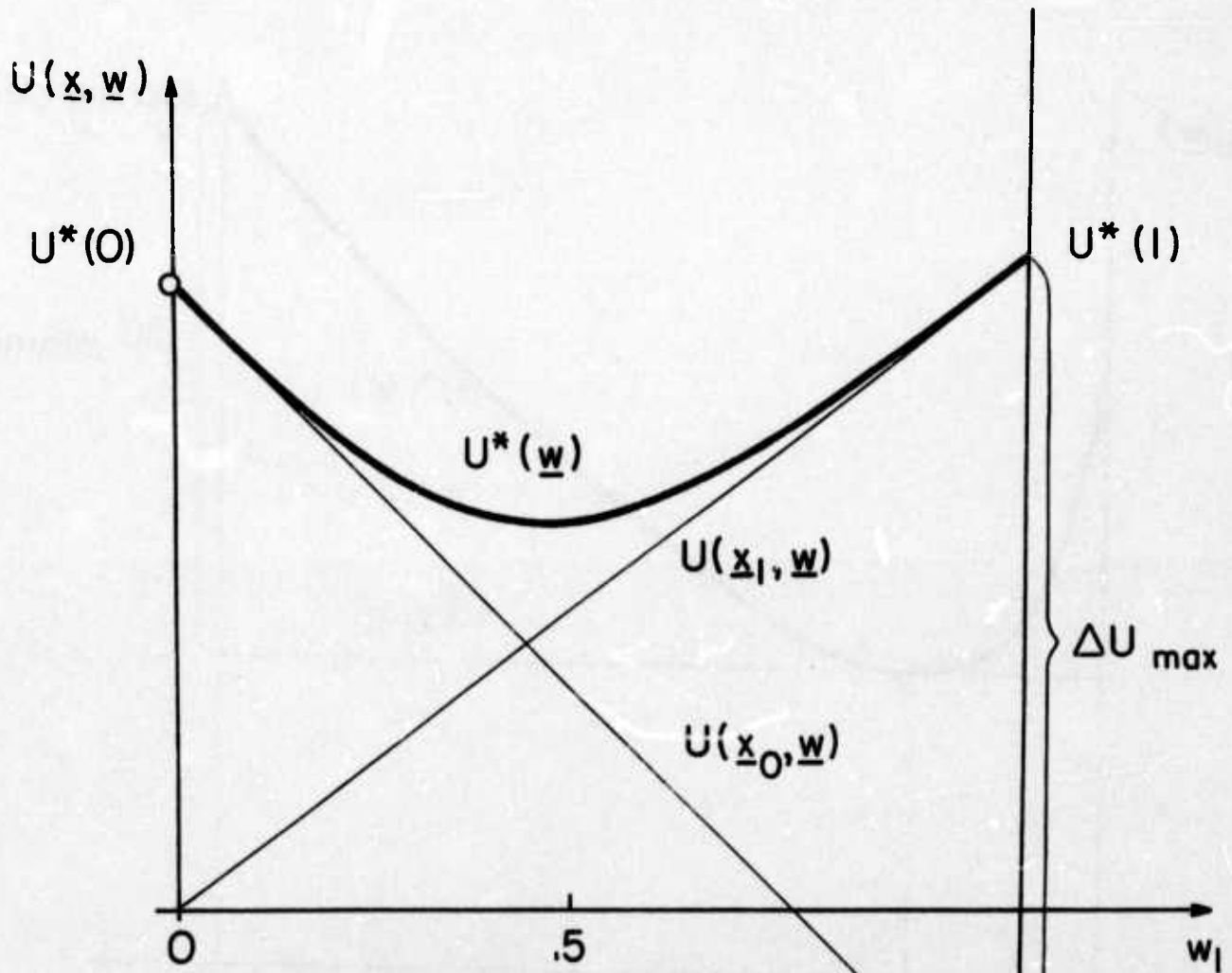


Figure 7

Graphical determination of the maximum possible loss  
under a minimax choice  $\underline{y}$  ( $\Delta U_{\text{minimax}}$ )

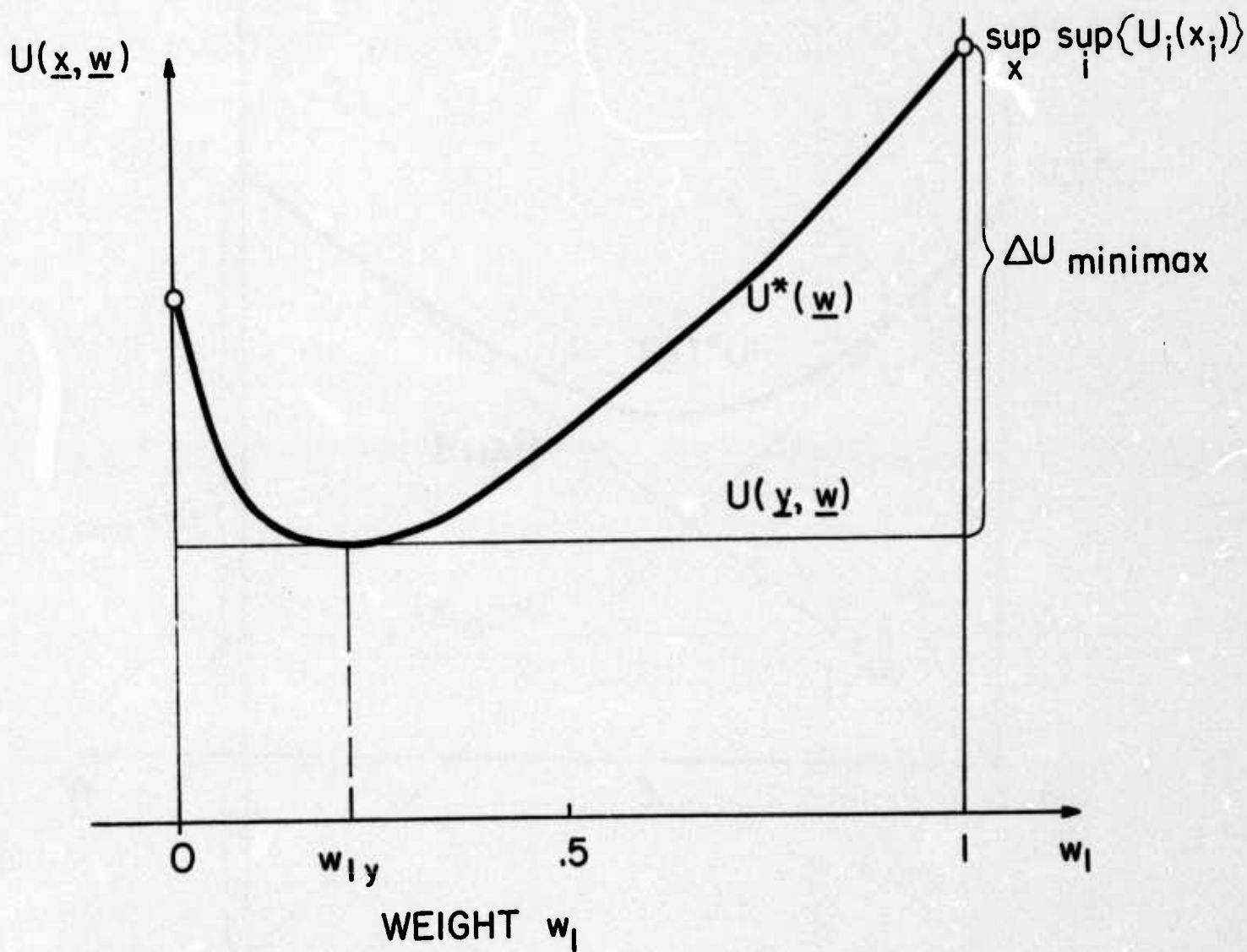


Figure 8

U and U\* functions in the signal detection example

( $m_1 = -.5; m_2 = +.5$ )

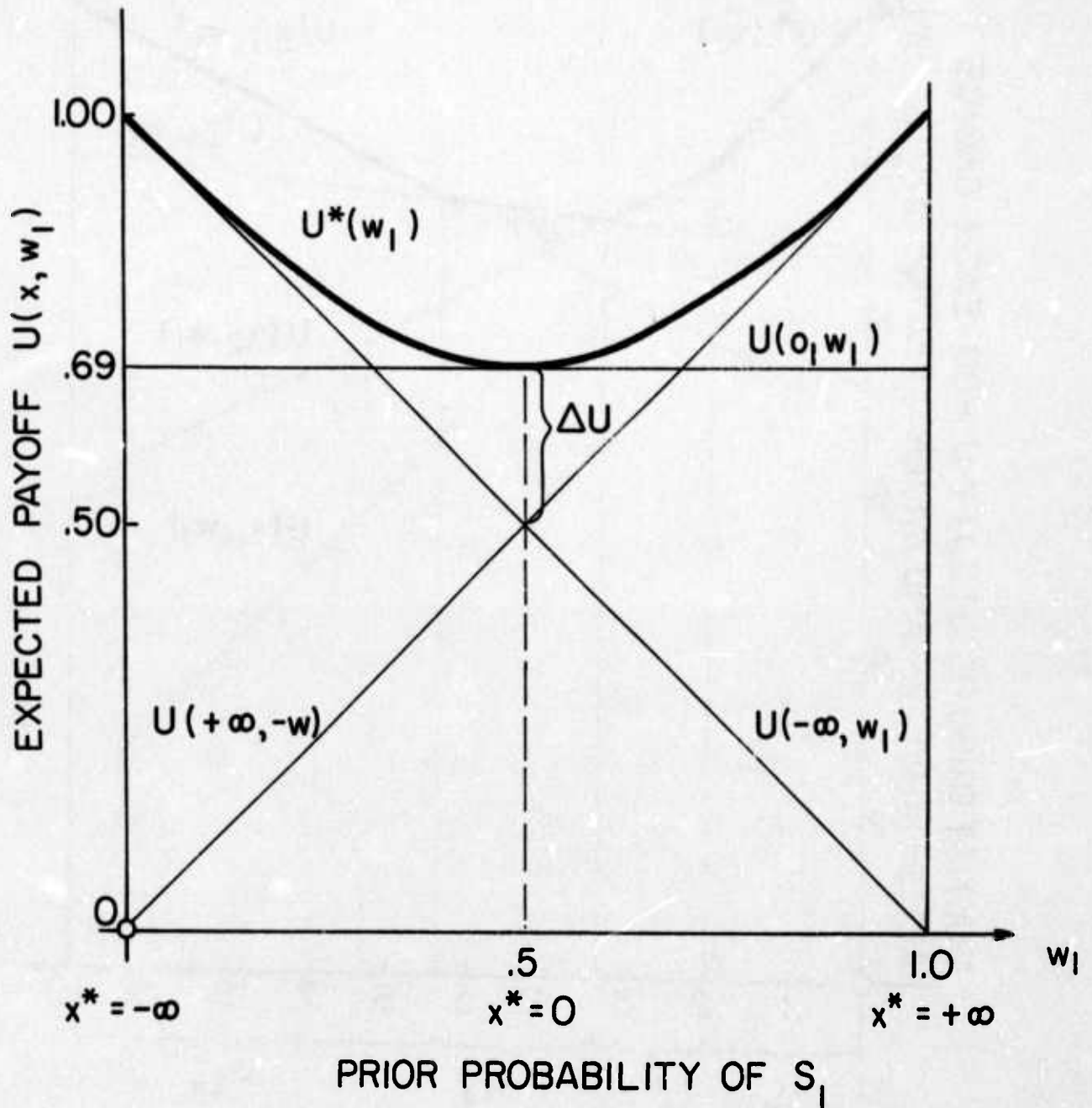


Figure 9

U and U\* functions in the multi-attribute example

