OPERATIONAL RESEARCH APPROACH TO DECISION MAKING

Oleg Pokrovsky

Abstract

The decision making (DM) problem is of great practical value in many areas of human activities. Most widely used DM methods are based on probabilistic approaches. The well-known Bayesian theorem for a conditional probability density function (PDF) is a background for such techniques. It is needed due to some uncertainty in many parameters entered in any model which describes the functioning of many real systems or objects. Uncertainty in our knowledge might be expressed in an alternative form. We offer to employ appropriate confidence intervals for model parameters instead of a relevant PDF. Thus one can formulate a prior uncertainty in model parameters by means of a set of linear constraints. The related cost or goal function should be defined at a corresponding set of parameters. That leads us to stating the problem in terms of operational research or mathematical linear programming. It is more convenient to formulate such optimization problems for discreet or Boolean variables. A review of relevant problem statements and numerical techniques are presented as well as many examples.

Keywords: *decision making, Bayesian theory, linear and integer programming, optimal design*

1 Introduction

Decision theory is a theory about decisions. The subject is not a very unified one. To the contrary, there are many different ways to theorize about decisions, and therefore also many different research traditions. This chapter attempts to reflect some of

Oleg Pokrovsky

Main Geophysical Observatory, Karbyshev str.7, St. Petersburg, 194021, Russian Federation, e-mail: pokrov@main.mgo.rssi.ru

the diversity of the subject. Its emphasis lies on the mathematical aspects of decision theory. Decision theory focuses on how we use our freedom. In the situations treated by decision theorists, there are options to choose between, and we choose in a non-random way. Our choices, in these situations, are goal-directed activities. Hence, decision theory is concerned with goal-directed behaviour in the presence of options. We do not decide continuously. In the history of almost any activity, there are periods in which most of the decision-making is made, and other periods in which most of the implementation takes place. Decision theory tries to throw light, in various ways, on the former type of period. Decision makers divide the decision process into the following five steps:

- Identification of the problem
- Obtaining necessary information
- Production of possible solutions
- Evaluation of such solutions
- Selection of a strategy for performance

The set of above issues is sequential in the sense that they divide decision processes into parts that always come in the same order or sequence. This approach might be criticized. Some empirical material indicates that the "stages" are performed in parallel rather than in sequence. A more realistic model should allow the various parts of the decision process to come in different order in different decisions.

2 Bayesian decision theory

Bayesian decision theory is based on the statistical inference in which evidence or observations are used to update or to newly infer the probability that a hypothesis may be true. The name "Bayesian" comes from the frequent use of Bayes' theorem in the inference process. Bayes' theorem was derived from the work of the Reverend Thomas Bayes. Bayesian inference uses aspects of the scientific method, which involves collecting evidence that is meant to be consistent or inconsistent with a given hypothesis. As evidence accumulates, the degree of belief in a hypothesis ought to change. With enough evidence, it should become very high or very low. Thus, proponents of Bayesian inference say that it can be used to discriminate between conflicting hypotheses: hypotheses with very high support should be accepted as true and those with very low support should be rejected as false. However, detractors say that this inference method may be biased due to initial beliefs that one needs to hold before any evidence is ever collected. Bayesian inference uses a numerical estimate of the degree of belief in a hypothesis before evidence has been observed and calculates a numerical estimate of the degree of belief in the hypothesis after evidence has been observed. Bayesian inference usually relies on degrees of belief, or subjective probabilities, in the induction process and does not necessarily claim to provide an objective method of induction. Nonetheless, some Bayesian statisticians

believe probabilities can have an objective value and therefore Bayesian inference can provide an objective method of induction.

$$P(H/E) = \frac{P(E/H)P(H)}{P(E)} \quad (1)$$

where:

- H represents a specific hypothesis, which may or may not be some null hypothesis.
- P(H) is called the prior probability of H that was inferred before new evidence, E, became available.
- P(E/H) is called the conditional probability of seeing the evidence E if the hypothesis H happens to be true. It is also called a likelihood function when it is considered as a function of H for fixed E.
- P(E) is called the marginal probability of E: the a priori probability of witnessing the new evidence E under all possible hypotheses. It can be calculated as the sum of the product of all probabilities of any complete set of mutually exclusive hypotheses and corresponding conditional probabilities:

$$P(E) = \sum_i P(E/H_i)P(H_i). \quad (2)$$

- P(H/E) is called the posterior probability of H given E.

The factor P(E/H) / P(E) represents the impact that the evidence has on the belief in the hypothesis. If it is likely that the evidence E would be observed when the hypothesis under consideration is true, but unlikely that E would have been the outcome of the observation, then this factor will be large. Multiplying the prior probability of the hypothesis by this factor would result in a larger posterior probability of the hypothesis given the evidence. Conversely, if it is unlikely that the evidence E would be observed if the hypothesis under consideration is true, but a priori likely that E would be observed, then the factor would reduce the posterior probability for H. Under Bayesian inference, Bayes' theorem therefore measures how much new evidence should alter a belief in a hypothesis.

Bayesian statisticians argue that even when people have very different prior subjective probabilities, new evidence from repeated observations will tend to bring their posterior subjective probabilities closer together. However, others argue that when people hold widely different prior subjective probabilities their posterior subjective probabilities may never converge even with repeated collection of evidence. These critics argue that worldviews, which are completely different initially, can remain completely different over time despite a large accumulation of evidence.

Thus, one applies Bayes theorem (see (1) and (2)), multiplying the prior by the likelihood function and then normalizing, to get the posterior probability distribution, which is the conditional distribution of the uncertain quantity given the data. A prior is often the purely subjective assessment of an experienced expert. Some will choose a conjugate prior when they can, to make calculation of the posterior distribution easier. In decision theory, a Bayes estimator is an estimator or decision

rule that maximizes the posterior expected value of a utility function or minimizes the posterior expected value of a loss function (also called posterior expected loss). Unfortunately, there are many decision making examples where Bayes theory fails due to difficulties in determining the prior probability distribution. Standard statistical practice ignores model uncertainty. Data analysts typically select a model from some class of models and then proceed as if the selected model had generated the data. This approach ignores the uncertainty in model selection, leading to over-confident inferences and decisions that are more risky than one thinks they are. Bayesian model averaging (BMA) provides a coherent mechanism for accounting for this model uncertainty. Several methods for implementing BMA have recently emerged (George, 1999, Raftery, 1996 a, b). Nonetheless, the BMA approach cannot solve the decision problem entirely (Weakliem, 1999).

3 Decision under severe uncertainty

It is common to make uncertain decisions (Ben-Haim, 2001). What can be done to make good (or at least the best possible) decisions under conditions of uncertainty? Info-gap robustness analysis evaluates each feasible decision by asking: how much deviation from an estimate of a parameter value, function, or set, is permitted and yet "guarantee" acceptable performance? In everyday terms, the "robustness" of a decision is set by the size of deviation from an estimate that still leads to performance within requirements when using that decision. It is sometimes difficult to judge how much robustness is needed or sufficient. However, according to info-gap theory, the ranking of feasible decisions in terms of their degree of robustness is independent of such judgments. To this end, the following questions must be addressed:

- What are the characteristics of decision problems that are subject to severe uncertainty?
- What difficulties arise in the modelling and solution of such problems?
- What type of robustness is sought?
- How does info-gap theory address these issues?

In what way is info-gap decision theory similar to and/or different from other theories for decision under uncertainty? Two important points need to be elucidated in this regard at the outset:

- Considering the severity of the uncertainty that info-gap was designed to tackle, it is essential to clarify the difficulties posed by severe uncertainty.
- Since info-gap is a non-probabilistic method that seeks to maximize robustness to uncertainty, it is imperative to compare it to the single most important "non-probabilistic" model in classical decision theory, namely Wald's maximin paradigm.

The maximin rule tells us to rank alternatives by their worst possible outcomes: we are to adopt the alternative the worst outcome of which is superior to the worst

outcome of the others. After all, this paradigm has dominated the scene in classical decision theory for well over sixty years. So, first let us clarify the assumptions that are implied by severe uncertainty:

1. A parameter λ, whose true value is subject to severe uncertainty.
2. A region of uncertainty Δ, where the true value of λ lies.
3. An estimate $\widetilde{\lambda}$ of the true value of λ.

Two remarks should be made with account to above assumption. First, The region of uncertainty is relatively large. Second, the estimate is a poor approximation of the true value of λ. Info-gap decision theory is radically different from all current theories of decision under uncertainty. The difference originates in the modeling of uncertainty as an information gap rather than as a probability. In general, info-gap's robustness model is a mathematical representation of a local worst-case analysis in the neighborhood of a given estimate of the true value of the parameter of interest. Under severe uncertainty the estimate is assumed to be a poor indication of the true value of the parameter and is likely to be substantially wrong. The fundamental question therefore is: given the severity of the uncertainty, the local nature of the analysis and the poor quality of the estimate, how meaningful and useful are the results generated by the analysis, and how sound is the methodology as a whole? The robust optimization literature (see Ben-Tal et al, 2006; Kouvelis, 1997) provides methods and techniques that take a global approach to robustness analysis. These methods directly address decision under severe uncertainty, and have been used for this purpose for more than thirty years. Wald's Maximin model is the main instrument used by these methods. The principal difference between the Maximin model employed by info-gap and the various Maximin models employed by robust optimization methods is in the manner in which the total region of uncertainty is incorporated in the robustness model. Info-gap takes a local approach that concentrates on the immediate neighborhood of the estimate. In sharp contrast, robust optimization methods set out to incorporate in the analysis the entire region of uncertainty, or at least an adequate representation thereof. In fact, some of these methods do not even use an estimate. The info-gap's robustness model is an instance of the generic Maximin model. Therefore, it is instructive to examine the mathematical programming (MP) formats of these generic models (Ecker and Kupferschmid, 1988; Thie 1988; Kouvelis and Yu, 1997).

4 Linear Programming

A Linear Programming (LP) problem is a special case of a Mathematical Programming problem (Dantzig, 1949; Kantorovich, 1966). From an analytical perspective, a mathematical program tries to identify an extreme (i.e., minimum or maximum) point of a function $f(x_1, x_2, ..., x_n)$, which furthermore satisfies a set of constraints, e.g., $g(x_1, x_2, ..., x_n) \geq b$. Linear programming is the specialization of mathematical programming to the case where both function f, to be called the objective function,

and the problem constraints g are linear. From an applications perspective, mathematical (and therefore, linear) programming is an optimization tool, which allows the rationalization of many managerial and/or technological decisions required by contemporary techno-socio-economic applications. An important factor for the applicability of the mathematical programming methodology in various application contexts is the computational tractability of the resulting analytical models. Under the advent of modern computing technology, this tractability requirement translates to the existence of effective and efficient algorithmic procedures able to provide a systematic and fast solution to these models. For Linear Programming problems, the Simplex algorithm provides a powerful computational tool, able to provide fast solutions to very large-scale applications, sometimes including hundreds of thousands of variables (i.e., decision factors). In fact, the Simplex algorithm was one of the first Mathematical Programming algorithms to be developed (Dantzig, 1949), and its subsequent successful implementation in a series of applications significantly contributed to the acceptance of the broader field of Operations Research as a scientific approach to decision making.

4.1 Illustrative Example

Let us consider a simple example of the MP problem formulation (Kantorovich, 1966). Assume that a company produces two types of products P_1 and P_2. Production of these products is supported by two workstations W_1 and W_2 , with each station visited by both product types. If workstation W_1 is dedicated completely to the production of product type P_1 , it can process 40 units per day, while if it is dedicated to the production of product P_2, it can process 60 units per day. Similarly, workstation W_2 can produce daily 50 units of product P_1 and 50 units of product P_2, assuming that it is dedicated completely to the production of the corresponding product. If the company's profit by disposing one unit of product P_1 is \$200 and that of disposing one unit of P_2 is \$400, and assuming that the company can dispose its entire production, how many units of each product should the company produce on a daily basis to maximize its profit?

First notice that this problem is an optimization problem. Our objective is to maximize the company's profit, which under the problem assumptions is equivalent to maximizing the company's daily profit. Furthermore, we are going to maximize the company profit by adjusting the levels of the daily production for the two items P_1 and P_2. Therefore, these daily production levels are the control/decision factors, the values of which we are asked to determine. In the analytical formulation of the problem the role of these factors is captured by modeling them as the problem *decision variables*:

- X_1 = number of units of product P_1 to be produced daily
- X_2 = number of units of product P_2 to be produced daily

In view of the above discussion, the problem objective can be expressed analytically as:

$$f(X_1, X_2) = 200X_1 + 400X_2. \quad (3)$$

Equation (3) will be called the objective function of the problem, and the coefficients 200 and 400, which multiply the decision variables in it, will be called the objective function coefficients.

Furthermore, any *decision* regarding the daily production levels for items P_1 and P_2, in order to be realizable in the company's operational context, must observe the production capacity of the two workstations W_1 and W_2. Hence, our next step in the problem formulation seeks to introduce these technological constraints. Let's focus first on the *constraint*, which expresses the finite production capacity of workstation W_1. Regarding this constraint, we know that one day's work dedicated to the production of item P_1 can result in 40 units of that item, while the same period dedicated to the production of item P_2 will provide 60 units of it. Assuming that production of one unit of product type $P_i(i = 1, 2)$, requires a constant amount of processing time $T_{1i}(i = 1, 2)$ at workstation W_1, it follows that: $T_{11} = \frac{1}{40}$ and $T_{12} = \frac{1}{60}$. Under the further assumption that the combined production of both items has no side-effects, i.e., does not impose any additional requirements for production capacity of the workstation (e.g., zero set-up times), the total capacity (in terms of time length) required for producing X_1 units of product P_1 and X_2 units of product P_2 is equal to $\frac{1}{40}X_1 + \frac{1}{60}X_2$. Hence, the technological constraint imposing the condition that our total daily processing requirements for workstation W_1 should not exceed its production capacity, is analytically expressed by:

$$\frac{1}{40}X_1 + \frac{1}{60}X_2 \leq 1. \quad (4)$$

Notice that in equation (4) time is measured in days.

Following the same line of reasoning (and under similar assumptions), the constraint expressing the finite processing capacity of workstation W_2 is given by:

$$\frac{1}{50}X_1 + \frac{1}{50}X_2 \leq 1. \quad (5)$$

Constraints (4) and (5) are known as the technological constraints of the problem. In particular, the coefficients of the variables $X_i(i = 1, 2)$, $\frac{1}{T_{ij}}(i, j = 1, 2)$, are known as the technological coefficients of the problem formulation, while the values on the right-hand-side of the two inequalities define the right-hand side vector of the constraints. Finally, to the above constraints we must add the requirement that any permissible value for variables $X_i(i = 1, 2)$ must be nonnegative since these values express production levels. These constraints are known as the variable sign restrictions. Combining equations (3) to (5), the analytical formulation of our problem is as follows:

$$max\{f(X_1, X_2)\} = max\{200X_1 + 400X_2\} \quad (6)$$

$$\frac{1}{40}X_1 + \frac{1}{60}X_2 \leq 1$$

$$\frac{1}{50}X_1 + \frac{1}{50}X_2 \leq 1$$

$$X_i \geq 0 (i = 1, 2).$$

4.2 The general "Linear Programming" formulation

Generalizing formulation (6), the general form for a Linear Programming problem is as follows (Gass, 1958):

Linear Objective Function (LOF) maximization:

$$max\{f(X_1, X_2, ..., X_n)\} = max\{\sum c_i X_i\} \quad (7)$$

under *Linear Constraints (LC)*:

$$\sum_j a_{ij} X_j \begin{array}{c} \leq \\ or \\ = \\ or \\ \geq \end{array} b_i (i = 1, ..., m). \quad (8)$$

The LC (8) might be used in important particular cases, when variables signs are prescribed:

$$(X_j \geq 0), or (X_j \leq 0). \quad (9)$$

We conclude our discussion on the general LP formulation by formally defining the solution search space and optimality. Specifically, we shall define as the feasible region of the LP of Equations (6) to (8), the entire set of vectors $\mathbf{X} = (X_1, ..., X_n)^T$ that satisfy the LC of (8) and the sign restrictions of (9). An optimal solution to the problem is any feasible vector that further satisfies the optimality requirements expressed by (7)-(9). Introducing integrality requirements for some of the variables in an LP formulation turns the problem to one belonging in the class of (Mixed) Integer Programming (MIP) or Integer Programming (IP).

4.3 Graphical LP's interpretation

In this section, we consider a solution approach for LP problems, which is based on a geometrical representation of the feasible region and the objective function

(Gass, 1958). In particular, the space to be considered is the n-dimensional space with each dimension defined by one of the LP variables (X_1, X_2). Thus we present an illustration for the 2-variable case.

We start our investigation regarding the geometrical representation of 2-var linear constraints by considering first constraints of the *equality type*, i.e.,

$$a_1 X_1 + a_2 X_2 = b. \quad (10)$$

Assuming $a_2 \neq 0$, this equation corresponds to a straight line with slope $s = \frac{-a_1}{a_2}$ and intercept $d = \frac{b}{a_2}$. In the special case $a_2 = 0$ the solution space (locus) of equation (10) is a straight line perpendicular to the X_1-axis, intersecting it at the point $(\frac{b}{a_1}; 0)$. Notice that the presence of an equality constraint restricts the dimensionality of the feasible solution space by one degree of freedom, i.e., it turns it from a planar area to a line segment.

Consider the *inequality constraint*:

$$a_1 X_1 + a_2 X_2 \begin{matrix} \leq \\ or \\ = \\ or \\ \geq \end{matrix} b. \quad (11)$$

The solution space of this constraint is one of the closed *half-planes* defined by equation (11). To show this, let us consider a point (X_1, X_2), which satisfies equation (11) as equality, and another point (X_1', X_2') for which equation (11) is also valid. For any such pair of points, it holds that:

$$a_1(X_1' - X_1) + a_2(X_2' - X_2) \begin{matrix} \leq \\ or \\ = \\ or \\ \geq \end{matrix} 0. \quad (12)$$

Let us consider the left side of (12) as the inner (dot) product of the two vectors $\mathbf{a} = (a_1, a_2)^T$ and $\Delta\mathbf{X} = ((X_1' - X_1), (X_2' - X_2))^T$. It is equal to $\left\|\Delta\mathbf{X}\right\| \left\|\mathbf{a}\right\| \cos(\Delta\mathbf{X}, \mathbf{a})$. In this case a line $a_1 X_1 + a_2 X_2 = b$ can be defined by the point (X_1, X_2) and the set of points (X_1', X_2') such that vector $\mathbf{a}$ is at right angles with vector $\Delta\mathbf{X}$. Furthermore, the set of points that satisfy the inequality parts of equation (12) have the vector forming an acute (obtuse) angle with vector $\mathbf{a}$, and therefore they are "above" ("below") the line. Hence, the set of points satisfying each of the two inequalities implied by equation (11) is given by one of the two half-planes the boundary of which is defined by the corresponding equality constraint. Figure 1 summarizes the above discussion.

An easy way to determine the half-plane depicting the solution space of a linear inequality is to draw the line depicting the solution space of the corresponding equality constraint and then test whether the point (0,0) satisfies the inequality. In
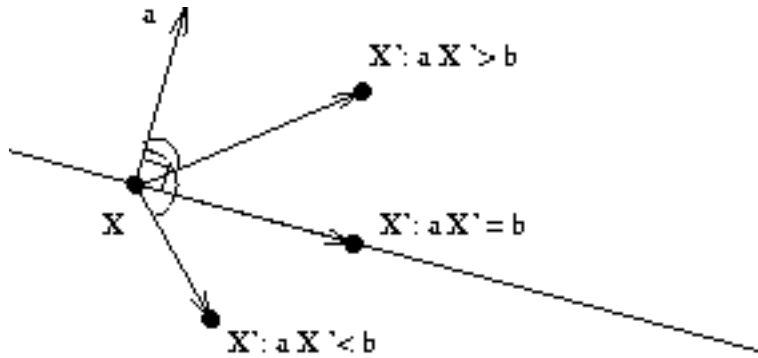
Fig. 1 Half-planes: the feasible region of a linear inequality

case of a positive answer, the solution space is the half-space containing the origin, otherwise, it is the other one.

From the above discussion, it follows that the feasible region for the prototype LP of equation (6) is the shaded area in the following figure:
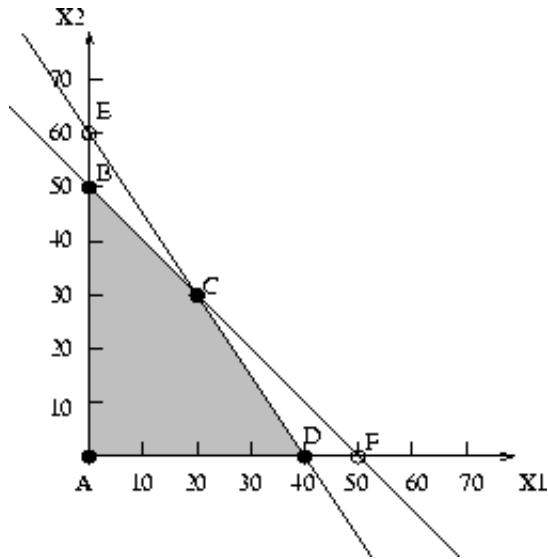

Fig. 2 The feasible region of the example LP considered in 3.1.

The next step is a maximization (minimization) of the objective function. The most typical way to represent a two-variable function $c_1 X_1 + c_2 X_2$ is to perceive it as a surface in an (orthogonal) three-dimensional space, where two of the dimensions correspond to the independent variables X_1 and X_2, while the third dimension provides the function value for any pair (X_1, X_2). However, in the context of our

discussion, we are interested in expressing the information contained in the two-var LP objective function $c_1X_1 + c_2X_2$ in the Cartesian plane defined by the two independent variables X_1 and X_2. For this purpose, we shall use the concept of contour plots. Contour plots depict a function by identifying the set of points (X_1, X_2) that correspond to a constant value of the function $(c_1X_1 + c_2X_2) = a$, for any given range of a's. The plot obtained for any fixed value of a is a contour of the function. Studying the structure of a contour is expected to identify some patterns that essentially depict some useful properties of the function. In the case of LP's, the linearity of the objective function implies that any contour of it will be of the type:

$$(c_1X_1 + c_2X_2) = a \quad (13)$$

i.e., a straight line. For a maximization (minimization) problem, this line will be called an isoprofit (isocost) line. Assuming that $c_2 \neq 0$, equation (13) can be rewritten as:

$$X_2 = -\frac{c_1}{c_2}X_1 + \frac{a}{c_2}$$

which implies that by changing the value of a , the resulting isoprofit/isocost lines have constant slope and varying intercept, i.e, they are parallel to each other (which makes sense, since by the definition of this concept, isoprofit/isocost lines cannot intersect). Hence, if we continuously increase a from some initial value a_o, the corresponding isoprofit lines can be obtained by "sliding" the isprofit line corresponding to $(c_1X_1 + c_2X_2) = a_o$ parallel to itself, in the direction of increasing or decreasing intercepts, depending on whether c_2 is positive or negative. The " sliding motion" suggests a way for identifying the optimal values for, let's say, a max LP problem. The underlying idea is to keep "sliding" the isoprofit line $(c_1X_1 + c_2X_2) = a_o$ in the direction of increasing a's until we cross the boundary of the LP feasible region. The implementation of this idea on the LP of equation (6) (see also fig.2) is depicted in figure 3.

From fig. 3, it follows that the optimal daily production levels for the protoype LP are given by the coordinates of the point corresponding to the intersection of line $\frac{1}{50}X_1 + \frac{1}{50}X_2 = 0$ with the X_2-axis, i.e., $X_1^{opt} = 0, X_2^{opt} = 50$. The maximal daily profit is $200*0 + 400*50 = 20,000\$$. Notice that the optimal point is one of the "corner" points of the feasible region depicted in Figure 3. Can you argue that for the geometry of the feasible region for 2-var LP's described above, if there is a bounded optimal solution, then there will be one which corresponds to one of the corner points? (This argument is developed for the broader context of n-var LP's in the next section.).

There are two fail options related to LP problem solution. First is absence of any solution, when the feasible region is empty. Consider again the original example (6), modified by the additional requirements (imposed by the company's marketing department) that the daily production of product X_1 must be at least 30 units, and that of product X_2 should exceed 20 units. These requirements introduce two new constraints into the problem formulation, i.e., $X_1 \geq 30, X_2 \geq 20$. Attempting to plot the feasible region for this new problem, we get figure 4, which indicates that
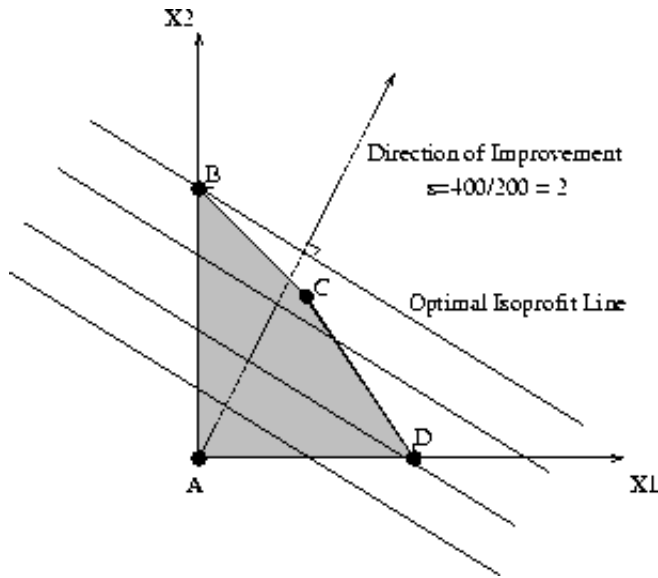
Fig. 3 Graphical solution of the example LP (6)

there are no points in the (X_1, X_2)-plane that satisfy all constraints, and therefore our problem is infeasible (over-constrained).
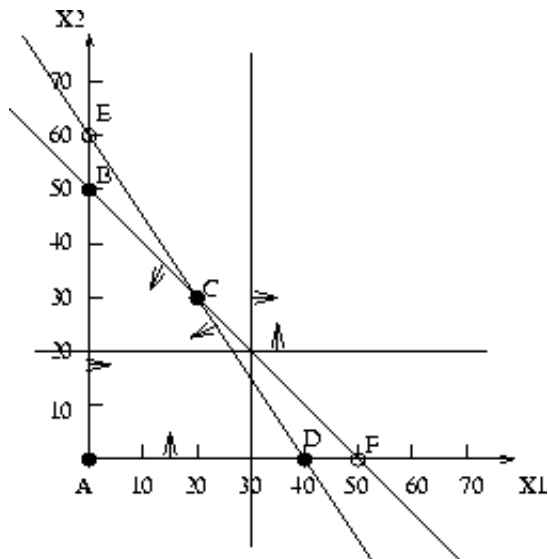

Fig. 4 An infeasible LP

A second particular option is an unbounded solution. In the LP's considered above, the feasible region (if not empty) was a bounded area of the plane. For this kind of problems it is obvious that all values of the LP objective function (and therefore the optimal) are bounded. Consider however the following modified LP problem:

$$max\{2X_1 - X_2\}$$

under constraints:

$$X_1 - X_2 < 1$$

$$2X_1 + X_2 > 6$$

$$X_1 \geq 0, X_2 \geq 0.$$

The feasible region and the direction of improvement for the isoprofit lines for this problem are given in figure 5. It is easy to see that the feasible region of this problem is unbounded, and furthermore, the orientation of the isoprofit lines is such that no matter how far we "slide" these lines in the direction of increasing the objective function, they will always share some points with the feasible region. Therefore, this is an example of a (2-var) LP whose objective function can take arbitrarily large values. Such an LP is characterized as unbounded. Notice, however, that even though an unbounded feasible region is a necessary condition for an LP to be unbounded, it is not sufficient; to convince yourself, try to graphically identify the optimal solution for the above LP in the case that the objective function is changed to:

$$max\{2X_1 - X_2\} = -X_2.$$

Summarizing the above discussion, we have shown that a 2-var LP can either:

- have a unique optimal solution which corresponds to a "corner" point of the feasible region, or
- have many optimal solutions that correspond to an entire "edge" of the feasible region, or
- be unbounded, or be infeasible.

5 Integer Programming

The use of integer variables in production when only integral quantities can be produced is the most obvious use of integer programs (Gomori, 1963; Korbut and Finkelstein, 1969). In this section, we will look at some less obvious ones. The text also goes through a number of them (some are repeated here).
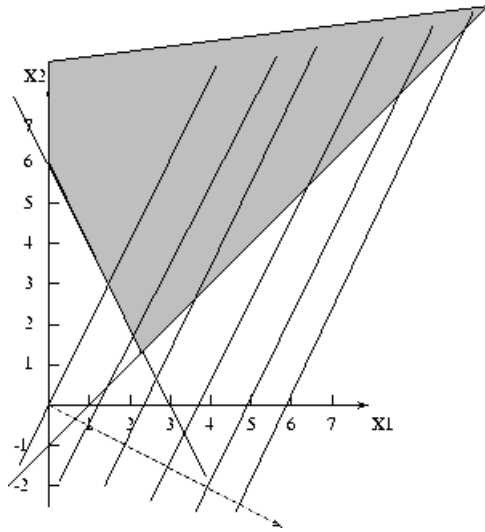
Fig. 5 An unbounded LP

5.1 Relationship to linear programming

Given is an Integer Program (IP):

$$max\{\mathbf{c^T} \cdot \mathbf{x}\}$$

subject to constraints:

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b}; \mathbf{x} \geq \mathbf{0}.$$

Since (LP) is less constrained than (IP), the following are immediate:

If (IP) is a minimization, the optimal objective value for (LP) is less than or equal to the optimal objective for (IP).

If (IP) is a maximization, the optimal objective value for (LP) is greater than or equal to that of (IP).

If (LP) is infeasible, then so is (IP).

If (LP) is optimized by integer variables, then that solution is feasible and optimal for (IP).

If the objective function coefficients are integer, then for minimization, the optimal objective for (IP) is greater than or equal to the "round up" of the optimal objective for (LP).

For maximization, the optimal objective for (IP) is less than or equal to the "round down" of the optimal objective for (LP). So solving (LP) does give some information: it gives a bound on the optimal value and, if we are lucky, may give the optimal solution to IP. We saw, however, that rounding the solution of LP will not in general

give the optimal solution of (IP). In fact, for some problems it is difficult to round and even get a feasible solution.

5.2 Capital budgeting

Let us consider one example of IP having a practical value (Gomori, 1963). Suppose we wish to invest $14,000. We have identified four investment opportunities. Investment 1 requires an investment of $5,000 and has a present value (a time-discounted value) of $8,000; investment 2 requires $7,000 and has a value of $11,000; investment 3 requires $4,000 and has a value of $6,000; and investment 4 requires $3,000 and has a value of $4,000. Into which investments should we place our money so as to maximize our total present value?

Our first step is to decide on our variables. This can be much more difficult in integer programming because there are very clever ways to use integrality restrictions. In this case, we will use a (0-1) variable $x_i (i = 1,..,4)$ for each investment. If x_i is 1 then we will make investment i. If it is 0, we will not make the investment. This leads to the 0-1 IP problem:

$$max\{8x_1 + 11x_2 + 6x_3 + 4x_4\}$$

subject to constraints:

$$5x_1 + 7x_2 + 4x_3 + 3x_4 \leq 14$$

$$x_i \in \{0;1\}, (i = 1,..,4).$$

Now, a straightforward decision suggests that investment 1 is the best choice. In fact, ignoring integrality constraints, the optimal linear programming solution is $(x_1 = 1; x_2 = 1; x_3 = 0.5; x_4 = 0)$ for a objective value of $22,000. Unfortunately, this solution is not integral. Rounding down to 0 gives a feasible solution with a value of $19,000. There is a better integer solution $(x_1 = 0; x_2 = 1; x_3 = 1; x_4 = 1)$, however, for an objective value of $21,000. This example shows that rounding does not necessarily give an optimal value.

5.3 Branch and Bound method

We discuss the branch and bound method by means of the simple IP example considered above. Our IP problem is as following:

$$max\{z\} = max\{8x_1 + 11x_2 + 6x_3 + 4x_4\}$$

subject to constraints:

$$5x_1 + 7x_2 + 4x_3 + 3x_4 \leq 14$$

$$x_i \in \{0;1\}, (i = 1,..,4).$$

The linear relaxation solution is $(x_1 = 1; x_2 = 1; x_3 = 0.5; x_4 = 0)$ with an objective function value of 22. We know that no integer solution will have value more than 22. Unfortunately, since x_3 is not integer, we do not have an integer solution yet. We want to force it to be integral. To do so, we branch on x_3, creating two new problems. In one, we will add the constraint x_3=0. In the other, we add the constraint x_3=1 . This is illustrated in figure 6.
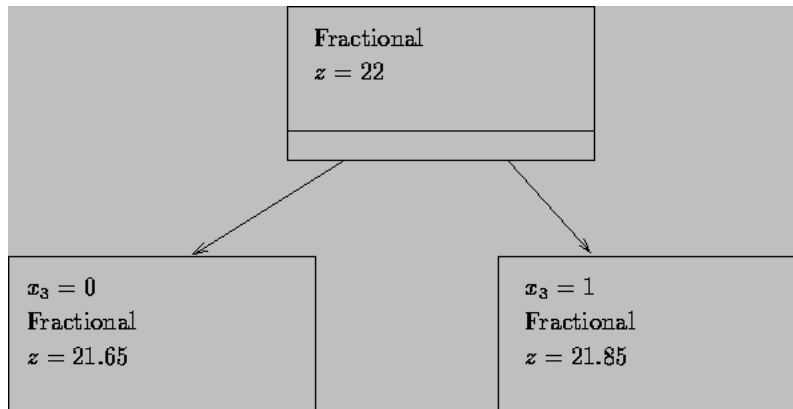

Fig. 6 First branching

Note that any optimal solution to the overall problem must be feasible to one of the subproblems. If we solve the LP by linear relaxations of the subproblems, we get the following solutions:

$$x_3 = 0; z = 21.65; for\, x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 0.667$$

$$x_3 = 1; z = 21.85; for\, x_1 = 1, x_2 = 0.714, x_3 = 1, x_4 = 0.$$

At this point we know that the optimal integer solution is no more than 21.85, but we still do not have any feasible integer solution. So, we will take a subproblem and branch on one of its variables. In general, we will choose the subproblem as follows:

- We will choose an active subproblem, which so far only means one we have not chosen before, and
- We will choose the subproblem with the highest solution value (for maximization; lowest for minimization).

In this case, we will choose the subproblem with $x_3=1$, and branch on x_2. After solving the resulting subproblems, we have the branch and bound tree in figure 7.
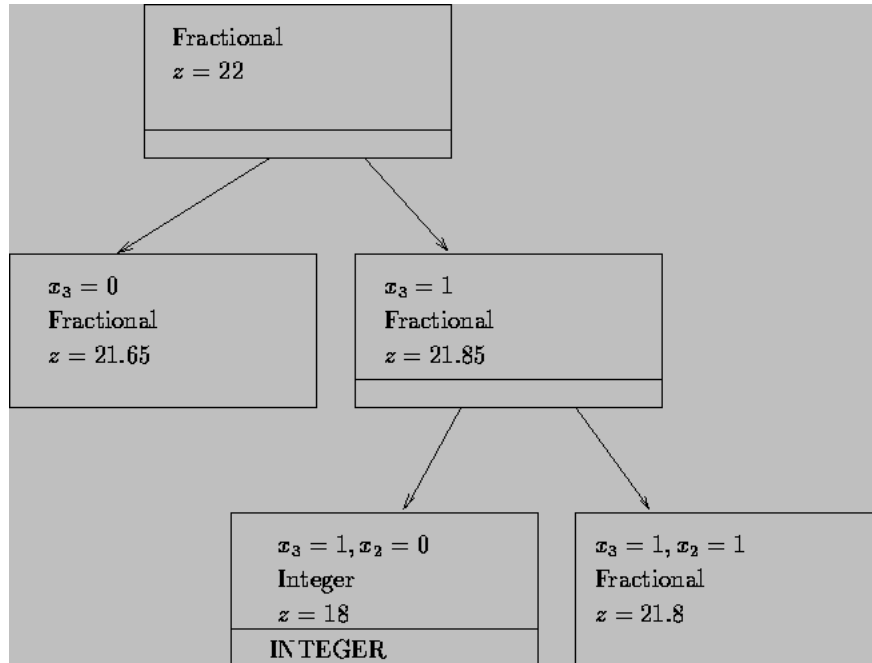

Fig. 7 Second branching

The solutions are:

$$x_3 = 1; z = 18; for x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1$$

$$x_3 = 1; z = 21.8; for x_1 = 0.6, x_2 = 1, x_3 = 1, x_4 = 0.$$

We now have a feasible integer solution $x_3 = 1, x_2 = 0$ with objective value 18. Furthermore, since the IP problem gave an integer solution, no further branching on that problem is necessary. It is not active due to the integrality of solution. There are still active subproblems that might give values more than 18. Using our rules, we will branch again to get figure 8.

The solutions are:

$$x_3 = 1, x_1 = 0, x_2 = 1; z = 21; for x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 1$$

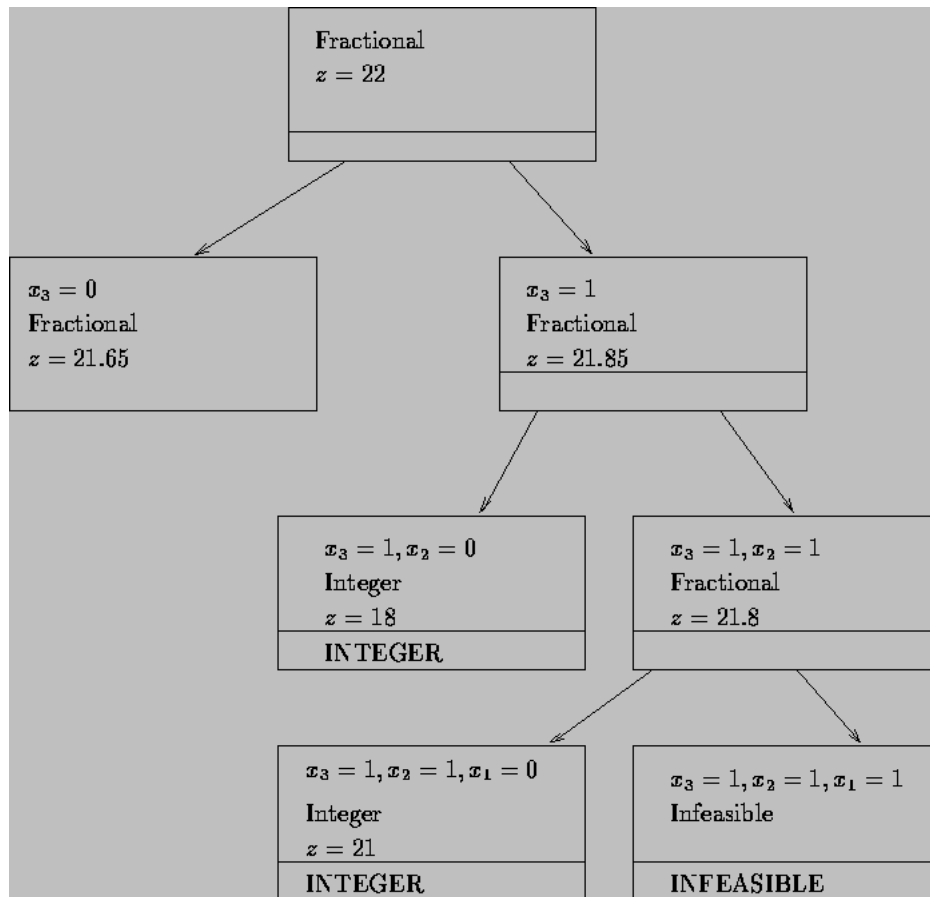$$x_3 = 1, x_1 = 1, x_2 = 1; infeasible.$$

Fig. 8 Third branching

Our best integer solution now has objective value 21. The subproblem that generates that is not active due to integrality of the solution. The other subproblem generated is not active due to infeasibility. There is still a subproblem that is active. It is the subproblem with solution value 21.65. By our "round-down" result, there is no better solution for this subproblem than 21. But we already have a solution with value 21. It is not useful to search for another such solution. We can fathom this subproblem based on the above bounding argument and mark it not active. There are no longer any active subproblems, so the optimal solution value is 21.

We have seen all parts of the branch and bound algorithm. The essence of the algorithm is as follows:

- Solve the linear relaxation of the problem. If the solution is integral, then we are done. Otherwise create two new subproblems by branching on a fractional variable.

- A subproblem is not active when any of the following occurs:

1. You used the subproblem to branch on,
2. All variables in the solution are integer,
3. The subproblem is infeasible,
4. You can fathom the subproblem by a bounding argument.

Choose an active subproblem and branch on a fractional variable. Repeat until there are no active subproblems.

6 The Integer Programming application to decision making

We considered several illustrative examples of possible applications of LP and IP. In this section we present several directions for practical application of IP in decision making. We will first discuss several examples of combinatorial optimization problems and their formulation as integer programs. Then we will review a general representation theory for integer programs that gives a formal measure of the expressiveness of this algebraic approach.

6.1 Main application areas

Formulating decision problems as integer or mixed integer programs is often considered an art form. However, there are a few basic principles which can be used by a novice to get started. As in all art forms though, principles can be violated to creative effect. We list below a number of example formulations, the first few of which may be viewed as principles for translating logical conditions into models.

6.1.1 Capacitated Plant Location Model

This model describes an optimal plan related to production and distribution of produced wares in accordance to demand sites. Let us introduce the following input parameters:

$i = \{1,...,m\}$- possible locations for plants

$j = \{1,...,n\}$−demand sites

k_i- a capacity of plant i; if opened

f_i- fixed cost of opening plant i

c_{ij}- per unit production cost at i plus transportation cost from plant i to site j

d_j- a demand at location j

Our task is to choose the plant locations so as to minimize total cost and meet all demands. This task might be formulated as the IP problem:

$$min\{\sum_j\sum_i c_{ij}x_{ij} + \sum_i f_i y_i\}$$

subject to constraints:

$$\sum_i x_{ij} \geq d_j; (j = 1, ..., n)$$

$$\sum_j x_{ij} \leq k_i y_i; (i = 1, ..., m)$$

$$x_{ij} \geq 0$$

$$y_i = \{0; 1\}.$$

to satisfy demands.

If the demand d_j is less than the capacity k_i for some "ij" combination, it is useful to add the constraint

$$x_{ij} \leq d_j y_i$$

to improve the quality of the linear programming relaxation.

6.1.2 Traveling Salesman Problem

A recurring theme in IP is that the same decision problem can be formulated in several different ways. Principles for sorting out the better ones have been the subject of some discourse (Korbut and Finkelstein, 1969). We now illustrate this with the well known traveling salesman problem. Given a complete directed graph with distance c_{ij} of arc (i; j), we are to find the minimum length tour beginning at node 1 and visiting each node of this graph exactly once before returning to the start node 1. This task might be formulated as the IP problem:

$$min\{\sum_j\sum_i c_{ij}x_{ij}\}$$

subject to constraints:

$$\sum_i x_{ij} = 1, (j = 1, ..., n)$$

$$\sum_j x_{ij} = 1, (i = 1, ..., n)$$

$$\sum_j\sum_i c_{ij}x_{ij} \geq 1$$

$$x_{ij} = \{0; 1\}.$$

6.1.3 Covering and Packing Problems

A wide variety of location and scheduling problems can be formulated as set covering or set packing or set partitioning problems. The three different types of covering, partitioning and packing problems can be succinctly stated as follows: Given

(a) a finite set of elements $M = \{1, ..., m\}$, and

(b) a family F of subsets of M with each member $F_j = \{1, ..., n\}$ having a profit (or cost) c_j associated with it,

find a collection, S, of the members of F that maximizes the profit (or minimizes the cost) while ensuring that every element of M is in:

(P1): at most one member of S (set packing problem)

(P2): at least one member of S (set covering problem)

(P3): exactly one member of S (set partitioning problem).

The three problems (P1), (P2) and (P3) can be formulated as integer linear programs as follows: Let A denotes the m*n matrix where

$$A_{ij} = \left\{ \begin{array}{c} 1, \, if \, element \, "i" \, belongs \, to \, F_j \\ 0, \, otherwise \end{array} \right\}.$$

The decision variables are $x_j (j = 1, ..., n)$, where

$$x_j = \left\{ \begin{array}{c} 1, \, if \, F_j \, is \, chosen \\ 0, \, otherwise \end{array} \right\}.$$

The set packing problem is

(P1)

$$max\{\mathbf{c^T} \cdot \mathbf{x}\}$$

subject to constraints:

$$\mathbf{A} \cdot \mathbf{x} \leq \mathbf{e}_m;$$

$$x_i = \{0; 1\}$$

where $\mathbf{e}_m$ is an m–dimensional column vector of "1"s. The set covering problem (P2) is (P1) with less than or equal to constraints replaced by greater than or equal to constraints and the objective is to minimize rather than maximize. The set partitioning problem (P3) is (P1) with the constraints written as equalities. The set partitioning problem can be converted to a set packing problem or a set covering problem (see (Korbut and Finkelstein, 1969)) using standard transformations. If the right hand side vector $\mathbf{e}_m$ is replaced by a non-negative integer vector $\mathbf{b}$, (P1) is referred to as the generalized set packing problem. The airline crew scheduling problem is a classic example of the set partitioning or the set covering problem. Each element of M corresponds to a flight segment. Each subset F_j corresponds to an acceptable set of flight segments of a crew. The problem is to cover, at minimum cost, each flight segment exactly once. This is a set partitioning problem.

6.2 Environmental application

6.2.1 Multi-User Consortium

Requirements for weather forecast products can vary significantly and are typically oriented to the needs of specific user groups. Nonetheless, in many respects the requirements are rather similar, such as a common need for information on basic variables such as temperature, humidity, and precipitation (mean, maximum, minimum). On other hand, it is hard to imagine that every user could provide their own forecast product because of substantial costs of both inputs and model development/maintenance. In the case of a specified forecast some additional observations might be required to increase prescribed reliability or probability. Therefore, it is more rational to select a set of a few forecast models and observing systems, which respond to the correct extent to an optimal set of requirements generated by a multi-user economical and mathematical model. A consortium of multi-users will get benefits of mathematically optimal decisions under minimal costs. User investments in a weather forecast system should be proportional to their expected benefits derived from the early warning of short-term weather fluctuations or extreme events. Under such circumstances a consortium of multi-users approach would be more likely to derive benefits from the mathematically optimal decisions for minimum investment. The meteorological community is interested in such an approach in order to reduce the number of observing programs and forecasting models (Pokrovsky, 2005; Pokrovsky, 2006).

6.2.2 Elementary Statement of Problem

Let us assume that there are *n users* of climate forecasting data with their *n benefits* of early warning: $c_i(i = 1,...,n)$ (i=1,.., n). These users are interested to forecast *m specific meteorological events* numerated as *j=1,...m*. The potential usefulness of them varies and is described by the matrix of coefficients $\mathbf{A} = \{a_{ij}\}$. Each magnitude a_{ij} can be considered as the expense of the *i-th user* for the *j-th meteorological event* delivered by some forecast model. The *minimum expected efficiency* for the *i-th user* is bounded by b_i^{min}. Let us introduce the decision maker variable:

$$x_i = \left\{ \begin{array}{c} 1, if\, user\text{``}i''adopts\, forecast\, data \\ 0,\, otherwise \end{array} \right\}.$$

Now we come to formulation of the optimization problem for $\{ x_i \}$:

$$max\{\sum_i c_i x_i\} \quad (14)$$

subject to constraints:

$$\sum_j a_{ij} x_j \geq b_i^{min}. \ (15)$$

Another interpretation of the coefficients and a more complex method to derive them is possible. A generalization to the forecast multi-model case is evident.
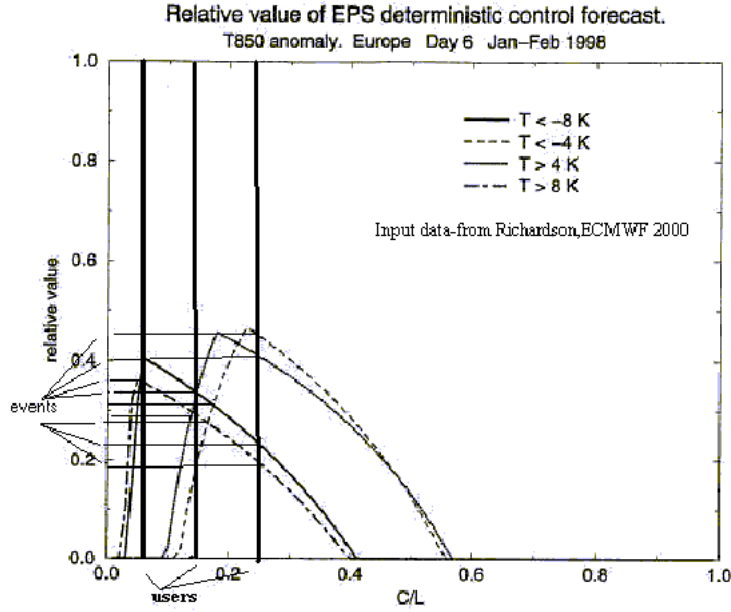

Fig. 9 Ensemble Prediction System forecast relative values (usefullness) responded to multi-user and multi-event case (Richardson, 2000)

6.2.3 Illustrative example

Let us consider multi-user decision making for many meteorological events. We used the European Center for Medium Range Weather Forecasting (ECMWF) Ensemble Prediction System (EPS) forecast for the T850 (air temperature field at the standard level of 850 mb) anomaly, Europe, Jan-Feb, 1998 (figure 9) (see details in Richardson, 2000) with n=3 (number of users), m=4 (number of meteorological events). The matrix of EPS forecast relative economic values are presented in table 1, the minimal efficiency for each user in table 2. In the case of equal importance of users we came to the optimal solution x_{opt} for (14) constrained by (15). This solution shows that the EPS forecasting system has prior importance for *user "2"*. The least contribution is related to *user "3"* . Let us now enhance the a priori importance of *user "3"* by changing values of the target function (1) from $\mathbf{c} = (1,1,1)^T$

to $\mathbf{c} = (0.5, 0.5, 1)^T$ (table 3 and table 4). Even in this case *user "3"* remains at second place after *user "1"*. It is interesting to note that the output for *user "1"* is its insensitivity with account to a priory weights.

Table 1 Matrix of constraints $\mathbf{A} = \{a_{ij}\}$

Users	(T_i-8K)	(T_i-4K)	($T_¿$+4K)	($T_¿$+8K)
1	0.40	0.36	0	0
2	0.32	0.29	0.32	0.19
3	0.22	0.19	0.41	0.46

Table 2 Constraint vector of minimal efficiencies -$\mathbf{b}_{min}$

Users	b_i^{min}
1	0.1
2	0.2
3	0.3

Table 3 Optimal decision x_{opt} in the case of priority user "3": $\mathbf{c} = (0.5, 0.5, 1)^T$

Users	x_{opt}
1	2.26
2	0.36
3	1.99

Table 4 Optimal decision x_{opt} in the case of priority user "3": $c = (0.5, 0.5, 1)^T$

Users	x_{opt}
1	2.26
2	0.36
3	1.99

7 Conclusion

An approach based on MP and IP finds a wide application area in many branches of economical sciences. It can be used in decision making related to multidimensional target functions constrained by many linear cost restrictions. This chapter indicates that similar problems arising in many important practical areas might be efficiently solved by the described approach.

References

1. Ben-Haim, Y. (2001) Information-Gap Theory: Decisions Under Severe Uncertainty, Academic Press, London.
2. Ben-Tal A., Ghaoui L. Al., and A. Nemirovski (2006) Mathematical Programming, Special issue on Robust Optimization, Volume 107(1-2).
3. Dantzig G.B. (1949) Programming of interdependent activities, Mathematical model. Econometrica, 17, 3, 200-211.
4. Ecker J.G. and Kupferschmid,M. (1988) Introduction to Operations Research, NY, Wiley.
5. Gass S.I. (1958) Linear Programming (methods and applications). NY-Toronto, London, "McGraw Hill".
6. George, E. I. (1999) Discussion of "Model averaging and model search strategies" by M. Clyde. In Bayesian Statistics 6 (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.), 157–185. Oxford University Press.
7. Gomory R.E. (1963) An algorithm for integer solution to linear programming. NY-Toronto, London, "McGraw Hill".
8. Kantorovich L.V. (1966) Mathematical models and methods of optimal economical planning. Novosibirsk, "Nauka", 256 p. (in Russian).
9. Korbut, A.A., and Yu.Yu. Finkelstein (1969) Discreet programming. Moscow, "Nauka", 302 p. (in Russian).
10. Kouvelis P. and G. Yu (1997) Robust Discrete Optimization and Its Applications, Kluwer.
11. Pokrovsky O.M. (2005) Development of integrated "climate forecast-multi-user" model to provide the optimal solutions for environment decision makers. - Proceedings of the Seventh International Conference for Oil and Gas Resources Development of the Russian Arctic and CIS Continental Shelf, St. Petersburg, 13-15 September, 2005, Publ. by AMAP, Oslo, Norway, September 2005, p. 661-663.
12. Pokrovsky O.M. (2006) Multi-user consortium approach to multi-model weather forecasting system based on integer programming techniques – Proceedings of the Second THORPEX International Scientific Symposium (Landshut, 3-8 December, 2006), WMO/TD, No. 1355, p.234-235.
13. Raftery, A. E. (1996a) Approximate Bayes factors and accounting from model uncertainty in generalised linear models. Biometrika 83, 251–266.
14. Raftery, A. E. (1996b) Hypothesis testing and model selection. In Markov Chain Monte Carlo in Practice (W. R. Gilks and D. Spiegelhalter, eds.) 163–188. Chapman and Hall, London.
15. Richardson R. (2000) Skill and relative economic value of the ECMWF Ensemble Prediction System. - Q. J. Roy. Met. Soc, 126, p.649-668.
16. Thie, P. (1988) An Introduction to Linear Programming and Game Theory, Wiley, NY.
17. Weakliem, D. L. (1999) A critique of the Bayesian information criterion for model selection. Sociological Methods and Research, 27, 359–297.