Adam Kasperski

Discrete Optimization with Interval Data

Studies in Fuzziness and Soft Computing, Volume 228

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 212. Elisabeth Rakus-Andersson
Fuzzy and Rough Techniques in Medical Diagnosis and Medication, 2007
ISBN 978-3-540-49707-3

Vol. 213. Peter Lucas, José A. Gàmez, Antonio Salmerón (Eds.)
Advances in Probabilistic Graphical Models, 2007
ISBN 978-3-540-68994-2

Vol. 214. Irina Georgescu
Fuzzy Choice Functions, 2007
ISBN 978-3-540-68997-3

Vol. 215. Paul P. Wang, Da Ruan, Etienne E. Kerre (Eds.)
Fuzzy Logic, 2007
ISBN 978-3-540-71257-2

Vol. 216. Rudolf Seising
The Fuzzification of Systems, 2007
ISBN 978-3-540-71794-2

Vol. 217. Masoud Nikravesh, Janusz Kacprzyk, Lofti A. Zadeh (Eds.)
Forging New Frontiers: Fuzzy Pioneers I, 2007
ISBN 978-3-540-73181-8

Vol. 218. Masoud Nikravesh, Janusz Kacprzyk, Lofti A. Zadeh (Eds.)
Forging New Frontiers: Fuzzy Pioneers II, 2007
ISBN 978-3-540-73184-9

Vol. 219. Roland R. Yager, Liping Liu (Eds.)
Classic Works of the Dempster-Shafer Theory of Belief Functions, 2007
ISBN 978-3-540-25381-5

Vol. 220. Humberto Bustince, Francisco Herrera, Javier Montero (Eds.)
Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, 2007
ISBN 978-3-540-73722-3

Vol. 221. Gleb Beliakov, Tomasa Calvo, Ana Pradera
Aggregation Functions: A Guide for Practitioners, 2007
ISBN 978-3-540-73720-9

Vol. 222. James J. Buckley, Leonard J. Jowers
Monte Carlo Methods in Fuzzy Optimization, 2008
ISBN 978-3-540-76289-8

Vol. 223. Oscar Castillo, Patricia Melin
Type-2 Fuzzy Logic: Theory and Applications, 2008
ISBN 978-3-540-76283-6

Vol. 224. Rafael Bello, Rafael Falcón, Witold Pedrycz, Janusz Kacprzyk (Eds.)
Contributions to Fuzzy and Rough Sets Theories and Their Applications, 2008
ISBN 978-3-540-76972-9

Vol. 225. Terry D. Clark, Jennifer M. Larson, John N. Mordeson, Joshua D. Potter, Mark J. Wierman
Applying Fuzzy Mathematics to Formal Models in Comparative Politics, 2008
ISBN 978-3-540-77460-0

Vol. 226. Bhanu Prasad (Ed.)
Soft Computing Applications in Industry, 2008
ISBN 978-3-540-77464-8

Vol. 227. Eugene Roventa, Tiberiu Spircu
Management of Knowledge Imperfection in Building Intelligent Systems, 2008
ISBN 978-3-540-77462-4

Vol. 228. Adam Kasperski
Discrete Optimization with Interval Data, 2008
ISBN 978-3-540-78483-8

Adam Kasperski

Discrete Optimization with Interval Data

Minmax Regret and Fuzzy Approach

Springer

Author

Dr. Adam Kasperski
Institute of Industrial Engineering and Management
Wroclaw University of Technology
Wybrzeze Wyspianskiego 27
Wroclaw, 50-370
Poland
Email: adam.kasperski@pwr.wroc.pl

ISBN 978-3-540-78483-8 e-ISBN 978-3-540-78484-5

DOI 10.1007/978-3-540-78484-5

Studies in Fuzziness and Soft Computing ISSN 1434-9922

Library of Congress Control Number: 2008922195

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Foreword

Operations research often solves deterministic optimization problems based on elegant and concise representations where all parameters are precisely known. In the face of uncertainty, probability theory is the traditional tool to be appealed for, and stochastic optimization is actually a significant sub-area in operations research. However, the systematic use of prescribed probability distributions so as to cope with imperfect data is partially unsatisfactory.

First, going from a deterministic to a stochastic formulation, a problem may become intractable. A good example is when going from deterministic to stochastic scheduling problems like PERT. From the inception of the PERT method in the 1950's, it was acknowledged that data concerning activity duration times is generally not perfectly known and the study of stochastic PERT was launched quite early. Even if the power of today's computers enables the stochastic PERT to be addressed to a large extent, still its solutions often require simplifying assumptions of some kind.

Another difficulty is that stochastic optimization problems produce solutions in the average. For instance, the criterion to be maximized is more often than not expected utility. This is not always a meaningful strategy. In the case when the underlying process is not repeated a lot of times, let alone being one-shot, it is not clear if this criterion is realistic, in particular if probability distributions are subjective. Expected utility was proposed as a rational criterion from first principles by Savage. In his view, the subjective probability distribution was basically an artefact useful to implement a certain ordering of solutions. However, subsequent research has shown that, in some situations, people having incomplete information on the problem at hand, or no information at all, consistently avoid making decisions by means of this criterion. There is a wealth of research papers in decision theory, in the last twenty years, that depart from the expected utility tradition.

And indeed, the use of single probability distributions in order to model imperfect data leads to another, more basic difficulty, namely it creates a confusion between values that are non-deterministic due to known variability, and values that are just unknown or ill-known while being possibly constant. The use of probability distributions is perfectly legitimate if sufficient statistics are available. In the face of mere ignorance, a probability distribution is either unreachable (when variability is present) or misleading. Subjective probability is a very rich concept that, due to its operational semantics, deliberately gives up the distinction between variability and ignorance.

In the recent past, a simpler idea has emerged, namely that partial ignorance on the value of a numerical parameter can be modelled by means of an interval. Intervals were used for a long time in physics for modelling measurement errors. But it never was the highlight in this area. Interval mathematics were further developed in the late sixties under the impulse given by Ramon Moore, motivated by the propagation of rounding errors made by digital computers. Since then, it has becomes a full-fledged approach to constraint-based problem solving and to the handling of uncertainty based on a best and worst case analysis. The advantage of this modeling is that it is not so much data-demanding as the probabilistic approach. The draw-back is that it is a cruder representation of uncertainty.

On this basis, there is no surprise that the interval-based approach became more popular in operations research in the 1990's (even if linear programming with interval coefficients was known for some time with the works of Ralph Steuer). Prior to devising algorithms adapted to this framework, a basic question to be addressed by optimization specialists is : what should be optimized if the criterion takes on interval values? Again decision theory has ready-made answers. One of the oldest such proposal is the maximin criterion: choose the solution that optimizes the criterion in the worst situation compatible the available incomplete information, i.e. maximize utility in the worst case. This criterion is very well known as being overpessimistic, as the worst situation may actually be very unlikely.

Interestingly, the same person who advocated expected utility, namely Leonard Savage, introduced the regret criterion, consisting of minimizing a loss function computed for each decision. The maximal regret associated with a decision is computed as the maximal difference between the figure of merit of the best decision that could have been taken had the underlying circumstances been known and the figure of merit of the considered decision in these circumstances. It turns out this is the most instrumental approach to state optimization problem under incomplete knowledge of data, so as to ensure a form of robustness to the selected solution.

This book considers solving combinatorial problems with incomplete data modelled by intervals using the regret criterion. The material is heavily based on the author's findings in his team in Wrocław, and focuses on some classical OR problems, such as knapsack, spanning tree, shortest paths, assignment, and minimal cut, and various sequencing problems in production research. Each OR researcher can then learn the regret minimization approach on her or his favourite problem. In this sense, the book makes an important contribution to the dissemination of more realistic approaches to combinatorial optimization, dropping the assumption of perfect knowledge of the data, without presupposing statistical information is available. The price paid is that, in general, the regret criterion leads to an increase in computational complexity, even if this increase is not so dramatic as in some stochastic approaches. From this point of view this book offers a detailed analysis to the complexity issues in each problem considered and proposes algorithms that cope with this difficulty.

Finally a significant merit of this book is to devote a chapter to the extension of the framework to the case where information is a little richer than what mere intervals allow to represents, namely when fuzzy intervals are allowed to model ill-known data. This step is very important for several reasons: first it introduces a formalism and a level of sophistication comparable with the probabilistic representations. It is the framework of possibility theory. Next, it is a direct extension, since a fuzzy interval is a collection of nested intervals. Another point is that the pessimism maximin criterion is tamed and becomes more sensible.

More often than not in the past, fuzzy versions of optimization problems were regarded with some doubt, because the obtained solutions were hard to interpret. This book highlights the fact that the fuzzy counterparts to operations research problems should be viewed as extensions of interval-valued formulations, and not of deterministic formulations. In other words if you are tempted to make your favorite method fuzzy because of imperfect information, try to state and solve the interval-valued problem first, and only then can you state and solve its fuzzy version on a safe ground. This book has the great merit to lay bare the importance of this methodology, thus opening a path to the rigorous, meaningful statement of fuzzy operations research problems in the setting of possibility theory, pursuing the pioneering works by the author's mentor, the late Stefan Chanas, and by Inuiguchi and Sakawa in Japan. No doubt this book is a very useful and original contribution to modern approaches in problem solving.

Toulouse
January 2008

Didier Dubois
Directeur de Recherche au CNRS

Preface

In operations research applications we are often faced with the problem of incomplete or uncertain data. This is caused by a lack of knowledge about the considered system or by a varying nature of the world. Uncertainty is a basic structural feature of the technological and business environment and it must be regarded as a part of the decision making process. In contrast, the classical mathematical models generally deal with the deterministic data. The uncertain parameters are often estimated to be the mean or the worst-case values over all possible realizations. The main weakness of the deterministic approach is that it does not take into account many possible realizations of the input data. In consequence, the obtained decision may be unacceptable in a varying and uncertain environment.

The *stochastic optimization* is a natural approach to model the uncertainty. It requires specifying a value of probability of every instance of the input data that may be realized. Then typically a decision is generated that minimizes or maximizes an expected performance measure. The stochastic approach has several disadvantages. It may be impossible or very expensive to estimate the probability distributions for the unknown data. The assumption of the distributional independence, which is often made to simplify the model, may be not justified. On the other hand the correlations between the parameters of the problem may be hard to identify. However, the most important failure of the stochastic optimization, pointed out by Kouvelis and Yu [87], is that it typically optimizes the expected system performance. But decision makers are reasonably more interested in hedging against the risk of poor (worst case) performance over all possible data realizations. This is particularly important for decisions that are encountered only once.

In this monograph we deal with an alternative approach to modeling the incomplete knowledge, which we refer to as the *robust optimization*. The idea and the framework of the robust optimization were described in a book by Kouvelis and Yu [87] and in the influential paper by Mulvey *et al* [107]. The robust approach is based on the concept of a *scenario*, which expresses a realization of the input data which may occur with some positive, but perhaps unknown probability. Before computing a solution the set of all scenarios, denoted by Γ, is identified. The aim is to find a solution that performs reasonably well under any scenario. Hence, contrary to the stochastic approach, in the robust optimization we minimize the worst-case system performance rather than the expected one.

There are two methods of determining the scenario set Γ. In the first case, called a *discrete scenario representation*, we explicitly list all scenarios, that is $\Gamma = \{S_1, S_2, \dots, S_K\}$, where $K \geq 1$. In the second case, called *interval scenario representation*, for every parameter in the problem there is a closed interval given and it is assumed that the value of the parameter may fall within this interval regardless of the values of the other parameters. In the interval scenario representation the scenario set Γ is the Cartesian product of all uncertainty intervals. Observe, that in this case Γ contains an infinite number of scenarios. The discrete scenario representation allows us to model the correlations among the input data. On the other hand, the interval representation assumes that the parameters in the problem are unrelated, that is the value of every parameter does not depend on the values of the remaining ones.

There are two types of uncertainty connected with the robust approach. The first appears when the feasibility of a solution depends on a particular scenario and the second is when the value of the cost (or the profit) of a given solution depends on a particular scenario. These two types of uncertainty were discussed by Mulvey *et al.* [107] and by by Kouvelis and Yu [87]. A solution that is "close" to optimal for any scenario is called *solution robust* and a solution that remains "almost" feasible for all scenarios is termed *model robust*. The notions of "close" and "almost" depend on the assumed optimization criterion. In this monograph we deal only with the first type of uncertainty, that is the one connected with the value of the objective.

In order to choose a solution for a given scenario set Γ several *robust criteria* were proposed in literature. Under *minmax criterion*, called also *absolute robust*, we seek a solution that minimizes the highest cost or maximizes the lowest profit over all scenarios. We thus assume that the worst will happen and we obtain in consequence a conservative, free of risk decision. Another criterion, first described by Savage [115], is that of *minmax regret*. The *maximal regret* (called also *robust deviation*) of a given solution expresses the maximal deviation of the cost (profit) of this solution from optimum over all scenarios. We seek a solution that minimizes the maximal regret. Using this criterion we obtain a solution which is less conservative and which minimizes the magnitude of missing opportunities. The maximal regret criterion is appropriate when the obtained decision is evaluated *ext post* and it is compared to the best decision that could have been made. For a deeper discussion on different robust criteria and the motivation of the robust approach we refer the reader to the book of Kuovelis and Yu [87] and to the paper of Mulvey *et al.* [107].

The main part of this monograph is devoted to the following class of robust optimization problems, namely *minmax regret discrete optimization problems*:

- A solution (decision) must be deriven from a finite set and it is either a subset or a permutation of a given, finite core set. Hence we consider the class of discrete optimization problems.
- The uncertainty appears only in the value of the objective. The feasibility of a solution does not depend on a scenario.
- The interval scenario representation is adopted. Thus for every uncertain parameter an interval of possible values, without a specific probability distribution, is given.
- In order to choose a solution the maximal regret criterion is applied. We thus seek a solution that minimizes the maximal regret.

The minmax regret approach to discrete optimization problems was discussed in book [87]. However, the major part of [87] is devoted to discrete scenario representation of uncertainty. Since the 1997s a lot of papers devoted to the interval uncertainty representation have appeared and the aim of this monograph is to present the state of the art in this field. This state of the art is still far from being complete and some open problems and questions are also addressed.

This monograph is divided into two parts. In the first part, composed of Chapters 1-11, we consider a class of discrete optimization problems in which the feasible solutions are formed by subsets of a given, finite core set of elements. A problem of this type is called a *combinatorial optimization problem.* A well known example is SHORTEST PATH in which we wish to find a path of the minimal total length between two distinguished nodes of a given graph. The uncertainty, modeled by closed intervals, appears in the weights associated to elements. These weights express the element costs, lengths, times *etc.* In Chapter 1 we provide the general formulation of the *minmax regret combinatorial optimization problem* and we show its general properties. In Chapter 1 we also discuss some related problems, in particular we discuss some other robustness criteria. Chapter 2 is devoted to a problem which is closely related to the minmax regret approach. In this chapter we introduce the notions of *possible* and *necessary* optimality of solutions and elements and we show some relationships between these concepts and the robust approach. The next two chapters, that is chapters 3 and 4 are devoted to general exact and approximation methods for solving the minmax regret combinatorial optimization problems. Among the exact methods are the mixed integer programming (MIP) formulation and the branch and bound algorithm. It turns out that a simple 2-approximation algorithm for this class of problems can also be designed. This algorithm, together with some of its extensions, are presented in Chapter 4. Chapters 5-9 are devoted to particular problems. We consider the very basic problems like MINIMUM SELECTING ITEMS, MINIMUM SPANNING TREE, SHORTEST PATH, MINIMUM ASSIGNMENT and MINIMUM CUT. For every problem we characterize its computational complexity and provide some exact and approximation algorithms.

In Chapter 10 we discuss a generalization of the minmax regret approach to combinatorial optimization problems. We show how the optimality evaluation and the solution concept can be generalized by extending the notion of the classical closed interval to a fuzzy one. We provide an interpretation of the fuzzy problem in terms of possibility theory. Namely, by introducing fuzzy intervals we can define a possibility distribution over scenario set. We show that all results obtained for interval problems can be applied to the fuzzy problems as well.

The second part of this monograph, composed of Chapters 12-16, is devoted to another class of discrete optimization problems, namely *sequencing problems*. In a sequencing problem every solution is a permutation of a given core set of elements, called *jobs*. Contrary to the combinatorial optimization problems considered in the first part, there may be several uncertain parameters associated with an element. In Chapter 12 we provide a general formulation of the *minmax regret sequencing problem*. In Chapters 13-15 we discuss three very basic problems.

In this monograph we consider only a particular class of discrete optimization problems. However, the minmax regret approach was also applied to the linear programming problem with interval coefficients in the objective function. For the results concerning this important problem we refer the reader to the papers of Inuiguchi and Sakawa [65], Mausser and Laguna [99, 100] and Averbakh and Lebedev [19].

In this monograph we do not discuss in detail the discrete scenario representation of uncertainty, which forms a separate and large class of problems. In Appendix A we only review briefly the known results and provide some new ones in this field. We compare them to the results known for the interval scenario representation.

Acknowledgments

I would like to dedicate this monograph to Professor Stefan Chanas, who is sadly no longer with us. Professor Chanas was not just my teacher; he was also my mentor who has played a great part in guiding me. Without his inspiration this monograph could not have been completed. I wish to express my gratitude to Professor Jacek Mercik and Professor Janusz Kacprzyk for their helpful suggestions and active interest in the publication of this monograph. I am indebted to Professor Didier Dubois who agreed to write the Foreword. I wish to express my thanks to all colleagues from my University, especially to Paweł Zieliński who has contributed to many results described in this book.

Wrocław
January 2008 Adam Kasperski

Contents

Part I: Minmax Regret Combinatorial Optimization Problems with Interval Data

1 Problem Formulation

We start this chapter by recalling the formulation of a deterministic combinatorial optimization problem, that is the one in which all the input parameters are precisely known. We also present some well known, special cases of this problem like minimum spanning tree, minimum selecting items, shortest path, minimum assignment and minimum cut. We introduce then the minmax regret combinatorial optimization problem with interval data, which can be viewed as a generalization of the deterministic problem where the imprecision is taken into account. We also discuss some closely related problems. In particular we consider some other robustness criteria which can be applied to the interval uncertainty representation.

1.1 Deterministic Combinatorial Optimization Problem

Let $E = \{e_1, \ldots, e_n\}$ be a finite ground set and let Φ be a set of subsets of E called the set of the *feasible solutions*. For every element $e \in E$ there is a nonnegative weight w_e given, which expresses a certain parameter like cost, length, time *etc.* associated with e. A *deterministic combinatorial optimization problem*, denoted by $\mathcal{P}$, consists in finding a feasible solution $X \in \Phi$ whose total weight is minimal, that is

$$\mathcal{P}: \min_{X \in \Phi} \sum_{e \in X} w_e. \tag{1.1}$$

The problem $\mathcal{P}$ can be alternatively represented as the following 0-1 programming problem with a linear objective function:

$$\min \textstyle\sum_{i=1}^{n} w_i x_i$$
$$(x_1, \ldots, x_n) \subset \{0, 1\}^n,$$

where x_i is a binary variable associated with element e_i and w_i is the weight of e_i. The set of feasible solutions is represented as a set of *characteristic vectors* of elements of Φ and it is typically described in a compact form by some set of constraints.

A. Kasperski: Discrete Optimization with Interval Data, STUDFUZZ 228, pp. 3–16, 2008.
springerlink.com

Formula (1.1) encompasses a wide range of problems. The following special versions of problem $\mathcal{P}$ play an important role in theory and applications:

- Minimum Spanning Tree: E is a set of edges of a given undirected graph $G = (V, E)$ and Φ consists of all spanning trees of G. For this problem we wish to find a spanning tree of G whose total weight is minimal. In a typical application we wish to connect a set of points (for instance terminals) using the least weight collection of edges.
- Minimum Selecting Items: E is a set of items and Φ consists of all subsets of E, whose cardinalities are exactly p, where $0 < p \leq n$ is a fixed integer. The problem consists in finding a subset of exactly p items whose total weight is minimal. This problem can be viewed as a basic resource allocation one or as the 0-1 Knapsack problem with unit capacities of the items.
- Shortest Path: E is a set of edges (arcs) of a given undirected (directed) graph G. Set Φ consists of all paths between two distinguished nodes s and t in G. We wish to find a path whose total weight (length) is minimal. This problem arises in plenty of applications. The most popular one is to design the shortest or the quickest tour between two cities. Hence V models a set of cities and E is a set of possible connections between them with specified lengths or traveling times.
- Minimum Assignment: E is a set of edges of a given bipartite graph $G = (V \cup W, E)$, $|V| = |W|$, and Φ is the set of all assignments (perfect matchings) in G. We seek an assignment whose total weight is minimal. This problem arises if we wish to pair some objects, for instance tasks and machines, to achieve the minimum total cost.
- Minimum Cut: E is a set of edges (arcs) of a given undirected (directed) graph G and s, t are two given nodes of G. Set Φ consists of all subsets of arcs that form $s - t$ cuts in G. An $s - t$ cut is a subset of arcs (edges) whose deletion disconnects s and t. For this problem we seek a minimum weighted $s - t$ cut. The Minimum Cut problem arises for instance in the analysis of traffic networks.

All the special versions of problem $\mathcal{P}$ described above are polynomially solvable and they have emerged as one of the major research topics in operations research since the 1950s. They arise frequently in practice and often lie at the heart of more complex problems.

In problem $\mathcal{P}$ we have assumed that we seek a solution that minimizes the total weight. Obviously, we can also consider a problem of determining a solution whose total weight is maximal. Hence we may consider problems Maximum Spanning Tree, Maximum Selecting Items, Longest Path, Maximum Assignment and Maximum Cut. Later in this monograph we will show that under a general assumption it is enough to consider only minimization problems.

Sometimes objective (1.1) is replaced with the *bottleneck* one, that is $\sum_{e\in X} w_e$ is replaced with $\max_{e\in X} w_e$. We will discuss the problems with the bottleneck objective in one of the next sections in this chapter.

Since $\mathcal{P}$ is a computational problem, we need to describe a way by which it is represented as the input to a computer algorithm. Obviously, listing all feasible solutions from Φ is very inefficient because Φ may contain up to $2^{|E|}$ solutions. Therefore, we will represent Φ in a compact form of a size polynomial in n. For instance, in terms of a polynomial algorithm, which decides whether a subset $X \subseteq E$ is a feasible solution (or it is a part of a feasible solution). For example in MINIMUM SPANNING TREE the input consists of a graph $G = (V, E)$ together with $|E|$ numbers denoting the weights of edges. We additionally provide an algorithm, which decides whether a given subset of edges does not contain a cycle.

1.1.1 Matroidal Problem

We distinguish now an important class of the deterministic combinatorial optimization problems namely *matroidal problems*. Let us recall that a *matroid* is a system $(E, \mathcal{I})$, where E is a finite set of elements and $\mathcal{I}$ is a set of subsets of E which fulfills the following two axioms:

(A1) if $A \subseteq B$ and $B \in \mathcal{I}$, then $A \in \mathcal{I}$,
(A2) if $A, B \in \mathcal{I}$ and $|A| < |B|$ then there is $e \in B \setminus A$ such that $A \cup \{e\} \in \mathcal{I}$.

The maximal under inclusion elements in $\mathcal{I}$ are called *bases* and the minimal under inclusion elements not in $\mathcal{I}$ are called *circuits*. It can be easily shown that all bases of a given matroid have the same cardinality. Perhaps the best known example of a matroid is a *graphic matroid* in which E is the set of edges of a given undirected graph and $\mathcal{I}$ consists of all acyclic subgraphs of G. A base of this matroid is a spanning tree of G. Another example is a *uniform matroid* in which E is a set of elements and $\mathcal{I}$ consists of all subsets of E whose cardinalities are less than or equal to a fixed integer p. A base in this matroid is a subset whose cardinality is exactly p

In the *matroidal combinatorial optimization problem* the set of feasible solutions Φ consists of all bases of a given matroid. In MINIMUM SPANNING TREE, Φ consists of the bases of a graphic matroid and in MINIMUM SELECTING ITEMS Φ consists of the bases of an uniform matroid. The remaining three problems considered in the previous section, that is SHORTEST PATH, MINIMUM ASSIGNMENT and MINIMUM CUT are not matroidal ones. What makes the matroidal problems very specific is that they are precisely the ones for which a simple `Greedy Algorithm` works. This algorithm is shown in Figure 1.1.

The `Greedy Algorithm` first arranges the elements in nondecreasing order of their weights. It picks then the elements one by one and tries to add it to the constructed solution X. The only case in which it excludes an element is when this element makes X infeasible. As the result it returns a solution, that is a base of a matroid X, whose total weight is minimal. The greedy algorithm for MINIMUM SPANNING TREE is known in literature as Kruskal's algorithm. For MINIMUM SELECTING ITEMS the greedy algorithm simply outputs p elements of the smallest weights. Due to their special structure, the matroidal problems will play an important role in this monograph.

```
Greedy Algorithm
```
Require: A matroidal combinatorial optimization problem $\mathcal{P}$
Ensure: An optimal solution $X \in \Phi$
1: Order elements so that $w_{e_1} \le w_{e_2} \le \cdots \le w_{e_n}$
2: $X \leftarrow \emptyset$
3: **for** $i \leftarrow 1$ **to** n **do**
4: **if** $X \cup \{e_i\} \in \mathcal{I}$ **then** $X \leftarrow X \cup \{e_i\}$
5: **end for**
6: **return** X

Fig. 1.1. The greedy algorithm that computes an optimal solution for a matroidal combinatorial optimization problem

1.2 Minmax Regret Combinatorial Optimization Problem with Interval Data

In practical applications the exact values of input data like costs, times, lengths *etc.*, are often not known in advance. It is caused by a lack of knowledge about a considered system or by the varying nature of the world. Perhaps, the simplest form of the uncertainty representation is to assume that the value of a given parameter may fall within a given range, independently on the values taken by the other parameters. Suppose that the values of the weights in problem $\mathcal{P}$ are only known to belong to the closed intervals $\tilde{w}_e = [\underline{w}_e, \overline{w}_e]$, where $\underline{w}_e \ge 0$ for all $e \in E$. If $\underline{w}_e = \overline{w}_e$, then the value of the weight of e is *precise* and in this case interval $\tilde{w}_e$ is called *degenerate*.

A particular realization of the weights $S = (w_e^S)_{e \in E}$ such that $w_e^S \in \tilde{w}_e$ for all $e \in E$ is called a *scenario*. Thus every scenario represents a certain state of the world, that is a configuration of the weights, which may occur with a positive, but perhaps unknown probability. We will denote by Γ the set of all scenarios, that is Γ is the Cartesian product of all the uncertainty intervals, namely $\Gamma = \times_{e \in E} \tilde{w}_e$. Among the scenarios we will distinguish the *extreme* ones, in which all the weights take the extreme values $\underline{w}_e$ or $\overline{w}_e$. Let $A \subseteq E$ be a given subset of E. We will use S_A^+ to denote the extreme scenario in which the elements $e \in A$ have weights $\overline{w}_e$ and all the other elements have weights $\underline{w}_e$. Similarly, we define scenario S_A^-, in which the elements $e \in A$ have weights $\underline{w}_e$ and all the other elements have weights $\overline{w}_e$. The particular scenarios S_A^+ and S_A^- will play a crucial role in further considerations.

The weight of a given solution $X \in \Phi$ under scenario $S \in \Gamma$ is defined as follows:

$$F(X, S) = \sum_{e \in X} w_e^S. \tag{1.2}$$

Let us denote by $F^*(S)$ the weight of the optimal solution under scenario S, that is

$$F^*(S) = \min_{X \in \Phi} F(X, S). \tag{1.3}$$

In order to obtain the value of $F^*(S)$ we must solve the deterministic combinatorial optimization problem $\mathcal{P}$ for the fixed scenario $S \in \Gamma$. The *maximal regret* of a given solution X is defined as follows:

$$Z(X) = \max_{S \in \Gamma}\{F(X,S) - F^*(S)\}. \tag{1.4}$$

It is clear that $Z(X) \geq 0$ for all solutions $X \in \Phi$. A scenario S that maximizes the right hand side of (1.4) is called the *worst case scenario for* X. We now prove the following proposition, which will be crucial in further considerations:

Proposition 1.1. *The scenario S_X^+ is the worst case scenario for solution X.*

Proof. Consider a scenario $S \in \Gamma$ and denote by Y the optimal solution under S, that is $F(Y,S) = F^*(S)$. It holds

$$F(X,S) - F^*(S) = \sum_{e \in X \setminus Y} w_e^S - \sum_{e \in Y \setminus X} w_e^S \leq \sum_{e \in X \setminus Y} \overline{w}_e - \sum_{e \in Y \setminus X} \underline{w}_e =$$
$$= F(X, S_X^+) - F(Y, S_X^+) \leq F(X, S_X^+) - F^*(S_X^+).$$

Hence S_X^+ maximizes the right hand side of (1.4) and S_X^+ is the worst case scenario for X. □

By Proposition 1.1 we can express the maximal regret of a given solution X in the following way:

$$Z(X) = F(X, S_X^+) - F^*(S_X^+) \tag{1.5}$$

or alternatively, making use of (1.2) and (1.5):

$$Z(X) = \max_{Y \in \Phi}\{\sum_{e \in X \setminus Y} \overline{w}_e - \sum_{e \in Y \setminus X} \underline{w}_e\} = \sum_{e \in X \setminus X^*} \overline{w}_e - \sum_{e \in X^* \setminus X} \underline{w}_e, \tag{1.6}$$

where X^* is the optimal solution under scenario S_X^+. The solution X^* is called the *worst case alternative for* X.

Equation (1.5) allows us to derive an important general conclusion. If the underlying deterministic combinatorial optimization problem $\mathcal{P}$ is polynomially solvable, then the maximal regret of a given solution X can be computed in polynomial time. It follows from the fact that in this case both scenario S_X^+ and the value of $F^*(S_X^+)$ can be obtained in polynomial time.

In this part of the monograph we focus on the following *minmax regret combinatorial optimization problem* $\mathcal{P}$:

$$\textsc{Minmax Regret } \mathcal{P}: \min_{X \in \Phi} Z(X).$$

Thus we seek a feasible solution for which the maximal regret is minimal. We will call the optimal solution to Minmax Regret $\mathcal{P}$ the *optimal robust solution*. Observe that Minmax Regret $\mathcal{P}$ is a natural generalization of problem $\mathcal{P}$ under uncertainty. If all the weight intervals are degenerate (in other words,