Department of Operations Research
University of Aarhus

June 2003

Classical and Non-Classical Stochastic Recourse Programs with Applications in Telecommunications

Morten Riis

PhD thesis

Classical and Non-Classical Stochastic Recourse Programs with Applications in Telecommunications

Morten Riis

PhD thesis

Thesis advisor: Kim Allan Andersen
Thesis committee: Jørgen Aase Nielsen, University of Aarhus
Rüdiger Schultz, Gerhard-Mercator-University of Duisburg
Maarten H. van der Vlerk, University of Groningen

Department of Operations Research
University of Aarhus

Morten Riis
Department of Operations Research
University of Aarhus, building 530
Ny Munkegade
DK - 8000 Århus C
Denmark
riis@imf.au.dk
http://home.imf.au.dk/riis

MSC2000. Primary: 90C15; secondary: 90B18, 90C06, 90C11, 90C31.
OR/MS subject classification. Programming: stochastic, integer; communications.

Preface

This thesis reflects the main part of the research I have done during my four years as a PhD student at the Department of Operations Research at the University of Aarhus. My research has been centered around stochastic (integer) programming problems with recourse, and hence it may be seen as a particular approach to the general field of decision-making under uncertainty.

Without engaging in a lengthy and technical account of the contents of the thesis, some explanation of the title may be appropriate at this point. In general, stochastic recourse programs are concerned with optimization problems for which the underlying decisions are organized in consecutive stages interspersed with the occurrences of some random events. Traditionally, the probability distribution of random parameters is assumed to be known, and the object of optimization is the expected value of some objective function (representing e.g. cost, revenue, profit, etc.). In this thesis, however, we will also be concerned with problems, fitting into the class of stochastic recourse programs, but for which a more appropriate object of optimization than the traditional one is found. Hence we will distinguish accordingly between classical stochastic recourse programs, i.e. the traditional expectation-based problems with known probability distributions of random parameters on the one hand, and what we choose to refer to as non-classical stochastic recourse programs on the other hand.

I have endeavored to pursue both theoretical and practical issues within the field of stochastic programming, and this is clearly reflected in the structure of the thesis. Following a short introduction to the field of stochastic programming in Chapter 1, the remainder of the thesis is organized in three major parts comprising a total of nine chapters. In the first part, comprised by Chapters 2 and 3, the reader is familiarized with basic well-known results and solution procedures for classical stochastic recourse programs. In the second part, comprised by Chapters 4 and 5, we consider two classes of non-classical stochastic recourse programs, and go through the analysis of theoretical properties and the elaboration of specialized solution procedures for the problems. Finally, in the third part, comprised by Chapters 6 through 10, five applications of stochastic programming, all within the context of communication networks, are presented.

The reader is assumed to be familiar with basic concepts from linear programming, integer programming, and to some extent probability theory. In Appendix A, though, we present some basic definitions and results from probability theory, that will be of importance throughout the thesis. The exposition is essentially self-contained, and it has been aspired to ensure that all chapters can be read independently of each other, even though some familiarity with the basic concepts and results presented in Chapters 1 to 3 is presumed throughout the rest of the thesis.

Acknowledgments

First of all I thank my supervisor Kim Allan Andersen for outstanding support and encouragement during the last four years. His insightful comments and suggestions always helped keeping this project on the right track.

During the late fall of 2000 I had the pleasure of enjoying a three month stay with Prof. Dr. Rüdiger Schultz at the Gerhard-Mercator University of Duisburg. During this stay I had the privilege to draw on his comprehensive and superior knowledge in the field of stochastic programming. My research has greatly benefited from the stimulating discussions we had during my stay, and since then he has been a recurring source of guidance and encouragement. For all this I am deeply indebted. Also, my gratitude is extended to Andreas Märkert, Stephan Tiedemann, Raymond Hemmecke, Markus Westphalen, and Dr. Ralf Gollmer for their friendly and helpful attitude and for making my stay in Duisburg socially as well as scientifically enjoyable.

At an early point of this project I decided to focus a major part of my research on applications of stochastic programming in telecommunications. The reason for my decision was the fact that telecommunications is a branch of trade undergoing a rapid and thriving development with the constant emergence of new technologies and services, and hence it naturally gives rise to decision problems calling for the application of stochastic programming. My decision was made, however, in spite of the fact that I knew nothing in particular about the world of telecommunications at that point in time. I am deeply grateful to Jørn Lodahl at SONOFON and Steen F. Møller at TDC for sharing their insight and expertise in this field with me, and for allowing me the opportunity to work with real-life applications and real-life data.

I would like to thank my colleagues at the Department of Operations Research for always providing a pleasant working atmosphere, and in particular credits are due to Anders J.V. Skriver for fruitful cooperation. Also, I thank Ali Khatam and Karsten N. Nielsen for countless "cigarette-breaks" (even though I do not smoke) and super friendship, and Randi Mosegaard for her invaluable linguistic support.

Last but not least I thank my family and in particular my wife to be, Bettina, for their love and support.

Århus, June 2, 2003

Morten Riis

Contents

Chapter 1

Introduction and Summary

The field of stochastic programming is concerned with optimization problems that are somehow infected by a certain degree of uncertainty with respect to the parameters and data of the underlying model. Thus, in stochastic programming, a decision problem under uncertainty is formalized by a mathematical model in which an appropriate objective for optimization is selected and uncertain parameters are represented as random variables. In this chapter we give a short introduction to the field of stochastic programming and in particular to the most common classes of stochastic programming models. The chapter is concluded with a summary of the remainder of the thesis.

1.1 The Foundation of Stochastic Programming

As suggested by the name, the foundation of stochastic programming is laid by modeling approaches and algorithmic techniques from mathematical programming. This distinguishes stochastic programming from other well-known approaches to the general field of decision-making under uncertainty, such as e.g. Markov decision processes, statistical decision theory, stochastic control theory etc.

With the invention of the simplex method in the late 1940s, the use of mathematical programming techniques, and in particular linear programming models, quickly gained widespread acceptance and popularity. It is a trivial observation, however, that uncertainty is almost always an inherent feature of the system to be controlled or analyzed. Moreover, in many applications of mathematical programming, the performance of the optimal solution provided by the model is seriously compromised if the actual state of nature turns out to differ from the specification of the model input. In general, these difficulties are much too profound and severe to be dealt with by means of ordinary sensitivity analysis or parametric analysis as the following small example suggests.

Example 1.1.1. Consider a small communications network with node set $V = \{1, 2, 3\}$ and edge set $E = \big\{\{1, 2\}, \{1, 3\}, \{2, 3\}\big\}$. Suppose that the network operator wishes to install capacity on links of this network so as to minimize total expected cost while meeting customer demand. Demand, however, is not known with certainty but is represented as a random variable. In particular, we assume that $D_{12} = D_{13} = \xi$ and $D_{23} = 1 - \xi$, where D_{ij} denotes demand for bandwidth on the link $\{i, j\} \in E$, and ξ is a uniformly distributed random variable, $\xi \sim U(0, 1)$. Suppose now that the network operator may install capacity on each link of the network at a unit cost of 1. If, however, insufficient capacity turns out to be available as the outcome of demand is observed, additional capacity must be rented from a competing network operator at a unit cost of 5. A naive approach to this problem is to solve the so-called *expected value problem*, replacing the random variable ξ by its expected value of $\frac{1}{2}$. Obviously, this approach results in the installment of $\frac{1}{2}$ unit of capacity on each link of the network. Recognizing the uncertainty of future demand, the network operator may choose to complement this solution with the results of a parametric analysis. In this case, the problem is easily solved for each possible outcome $\xi = t$, the optimal solution being $x_t = (t, t, 1-t)$ for $0 \leq t \leq 1$. The corresponding expected cost is

$$C(x_t) = t + t + (1-t) + 5\int_0^t (t-\xi)d\xi + 10\int_t^1 (\xi - t)d\xi = \frac{15}{2}t^2 - 9t + 6.$$

We see that the best solution with respect to total expected cost generated by the parametric analysis is $x_{\frac{3}{5}} = \left(\frac{3}{5}, \frac{3}{5}, \frac{2}{5}\right)$ with expected cost of 3.3, whereas the solution of the expected value problem yields total expected cost of 3.375. These solutions are, however, inferior to the solution $x = \left(\frac{4}{5}, \frac{4}{5}, \frac{4}{5}\right)$ that yields total expected cost of only 2.7.

As illustrated by Example 1.1.1, and discussed more thoroughly by e.g. Wallace [161], sensitivity analysis or parametric analysis does not provide an adequate framework for decision-making under uncertainty. Instead, what is required, is a model that somehow explicitly takes into account the uncertainty that infects model input. First steps in this direction were made with the pioneering work of Beale [11] and Dantzig [37] in the mid 1950s, introducing the class of so-called two-stage stochastic programs with recourse. Also, another line of research was initiated a few years later as Charnes and Cooper [35] introduced the class of so-called chance-constrained stochastic programming models. In the following section we give a brief introduction to these model classes.

1.2 Stochastic Programming Models

In this section we present some basic modeling approaches in stochastic programming and give a few examples of how different stochastic programming models may arise from some particular decision problem under uncertainty. To keep the exposition in line with the remainder of the thesis, we take as starting point a (mixed-integer) linear programming problem in which some parameters are not known with certainty at the point of decision. For a more general and comprehensive introduction to stochastic programming models, including among other things non-linear formulations of the objective function and the constraints, we refer to the textbooks by Birge and Louveaux [22], Kall and Wallace [67] and Prékopa [107]. Here we will consider the following random (mixed-integer) linear programming problem,

$$\begin{aligned} \min\ & cx, \\ \text{s.t. } & Ax \geq b, \\ & "T(\omega)x \geq h(\omega)", \\ & x \in X. \end{aligned} \tag{1.2.1}$$

Here $c \in \mathbb{R}^{n_1}$ and $b \in \mathbb{R}^{m_1}$ are known vectors, A is a known $m_1 \times n_1$-matrix, and $X \subseteq \mathbb{R}^{n_1}$ is some subset that may or may not contain integrality restrictions on some or all of the variables $x \in \mathbb{R}^{n_1}$. Uncertainty is reflected in the model by the fact that the coefficient matrix T and the right-hand side h of the second group of constraints are dependent on the outcome of some random event ω. Denoting by $\mathbb{R}^{m \times n}$ the space of real $m \times n$-matrices, it is assumed that $T : \Omega \mapsto \mathbb{R}^{m_2 \times n_1}$ and $h : \Omega \mapsto \mathbb{R}^{m_2}$ are measurable mappings defined on some probability space $(\Omega, \mathcal{F}, P)$. Note that transposes have been eliminated for simplicity throughout the thesis.

It is important to note at this point that we are seeking a here-and-now decision x, that must be based solely on the information available at the point of decision. Hence, in particular, the decision cannot be based on the actual outcome of the random event ω, since the only information presently available about this future event is conveyed through the distribution P. In other words, we say that the decision x should be *non-anticipative*. Obviously this means that problem (1.2.1) is not well-defined since the second group of constraints do not make sense when the decision x must be made before the outcome of the random event ω is known (and hence the quotation marks in (1.2.1)). Thus, the challenge of the model builder is to replace problem (1.2.1) by a well-defined problem, producing a non-anticipative solution that is in some sense optimal. The notion of optimality, though, is no longer obvious since the choice of an objective for minimization as well as the formulation of the constraints depend to some extent on the preferences of the decision-maker, and in particular on the attitude of the decision-maker toward risk. In the subsequent sections we present the most common examples of stochastic programming models that may arise from problem (1.2.1).

Remark 1.2.1. It is customary in stochastic programming to assume that the probability distribution of random parameters is known. This may seem like a rather strong assumption, and quite surely it will not be truly conformed with in many practical situations. This is not to say, however, that the assumption is not reasonable and justifiable. Thus, given the fact that uncertainty is an inherent feature of some particular decision problem, the decision-maker may choose to incorporate the uncertainty into some stochastic programming model, or he may simply choose to ignore it. In the latter case, though, values of the uncertain parameters must still be specified, and hence this approach really comes down to specifying a probability distribution where all mass is put into a single point. Therefore, it seems obvious that a more accurate and appropriate model is obtained if some effort is at least put into the estimation of the probability distribution of random parameters. In Chapter 5 we will discuss a particular approach to stochastic programs where the probability distribution is unknown, in the sense that the only information available pertains to estimates of certain moments or other distributional characteristics of the random parameters.

Remark 1.2.2. In this thesis we only consider so-called *anticipatory* problems, where decisions must be made without certain knowledge about future outcomes of random variables as explained above. In the class of *distribution problems*, on the other hand, decisions are made after uncertainty has been revealed. Hence the problem (1.2.1) is viewed as an ordinary linear program for each outcome of the random event ω, and the effort lies in determining distributional characteristics (i.e. mean values and higher moments, quantiles etc.) of the corresponding optimal solutions and their objective values.

Thus the distribution problem may in this case be seen as a generalization of sensitivity analysis or parametric analysis in linear programming.

1.2.1 Chance-Constrained Programming

Consider problem (1.2.1) and assume that the decision-maker requires the second group of constraints to hold with a probability of at least α, where $\alpha \in [0, 1]$. This leads to the following formulation of the problem,

$$\begin{aligned} \min\ & cx, && (1.2.2a)\\ \text{s.t. } & Ax \geq b, && (1.2.2b)\\ & P\big(T(\omega)x \geq h(\omega)\big) \geq \alpha, && (1.2.2c)\\ & x \in X. && (1.2.2d) \end{aligned}$$

Problem (1.2.2) is an example of a chance-constrained stochastic program, and (1.2.2c) is referred to as a joint chance-constraint or a joint probabilistic constraint. Clearly this approach allows for some generalizations. In particular, we may consider the separate chance-constraints

$$P\big(T_k(\omega)x \geq h_k(\omega)\big) \geq \alpha_k, \quad k = 1, \dots, m_2,$$

where T_k is the kth row of T, h_k is the kth component of h, and α_k is the probability with which the kth constraint is required to hold. Also, one may consider a combination of these approaches, clustering the individual constraints into a number of groups that must each hold jointly with some prescribed probability.

We will not consider chance-constrained problems any further in this thesis. For an overview of the research in this area we refer to Prékopa [107] and references therein.

1.2.2 Two-Stage Stochastic Programs with Recourse

Throughout this thesis we will primarily be concerned with two-stage stochastic recourse programs. These are problems in which the decisions can be divided in two groups — a group of first-stage decisions that must be made without certain knowledge about some random parameters of the model, and a group of second-stage decisions that can be taken after uncertainty has been revealed. For example, consider problem (1.2.1) and assume that a temporary violation of the second group of constraints is allowed. Feasibility must be restored, however, through some corrective or *recourse* actions $y(\omega)$ that can be taken after the actual outcome of ω has been observed. Assuming that the decision-maker seeks to minimize the sum of direct cost and expected recourse cost (i.e. the classical approach in stochastic programming), the problem may now be given the following formulation,

$$\begin{aligned} \min\ & cx + \mathbb{E}[q(\omega)y(\omega)], && (1.2.3a)\\ \text{s.t. } & Ax \geq b, && (1.2.3b)\\ & T(\omega)x + W(\omega)y(\omega) \geq h(\omega), \quad P-a.s. && (1.2.3c)\\ & x \in X, y(\omega) \in Y, \quad P-a.s. && (1.2.3d) \end{aligned}$$

Here we assume that $q : \Omega \mapsto \mathbb{R}^{n_2}$ and $W : \Omega \mapsto \mathbb{R}^{m_2 \times n_2}$ are measurable mappings defined on the probability space $(\Omega, \mathcal{F}, P)$ cf. our assumptions on h and T. The dependency of y on ω, on the other hand, is of a completely different nature, merely indicating that the recourse actions typically differ under different realizations of the random event. The constraints (1.2.3c) and (1.2.3d) are assumed to hold P-almost surely, i.e. for all $\omega \in \Omega \setminus N$, where $N \subseteq \Omega$ is some set such that $P(N) = 0$. Finally, $Y \subseteq \mathbb{R}^{n_2}$ is some subset that may or may not contain integrality restrictions on some or all of the recourse variables. Throughout this thesis we will refer to c as the first-stage cost, to q as the second-stage cost, to T as the technology matrix, to W as the recourse matrix, and to h as the second-stage right-hand side.

Remark 1.2.3. In formulating problem (1.2.3) we chose to see the first-stage variables as the main decisions to be made, whereas the second-stage variables were interpreted simply as corrective actions, providing a means of penalizing infeasibilities arising from the first-stage decision. It is important to note, however, that the two-stage stochastic recourse model applies to a far wider range of problems. Thus, the classification of decisions into stages may in fact be an inherent feature of an overall problem, implied simply by the timing of decisions — some decisions may have to be made immediately, whereas others can be postponed until uncertainty has been disclosed, or at least until additional information on uncertain parameters is available. Still, we will use the terms second-stage variables and recourse variables interchangeably, whether the variables represent simple corrective actions or actual decisions that are an immanent part of the problem.

As mentioned above, the two-stage stochastic program with recourse represents a simple dynamic decision process. The first-stage decision must be made without certain knowledge about random parameters, and must be chosen so as to minimize the sum of direct cost and the expected value of future cost. The future cost, on the other hand, is determined when an optimal second-stage decision is made after observation of the outcome of the random parameters. Thus the decision process may be summarized as

$$\text{decision on } x \rightarrow \text{observation of } q(\omega),\ h(\omega),\ T(\omega), \text{ and } W(\omega) \rightarrow \text{decision on } y.$$

The dynamics of the decision process is clearly illustrated by the following alternative dynamic programming formulation of the problem,

$$\min\bigl\{cx + \mathcal{Q}(x) \mid Ax \geq b,\ x \in X\bigr\}, \tag{1.2.4}$$

where the so-called expected recourse function $\mathcal{Q}$ is given by

$$\mathcal{Q}(x) = \mathbb{E}[\Phi(x,\omega)] = \int_{\Omega} \Phi(x,\omega) P(d\omega), \tag{1.2.5}$$

and the second-stage value-function Φ is given by

$$\Phi(x,\omega) = \min\bigl\{q(\omega)y \mid W(\omega)y \geq h(\omega) - T(\omega)x,\ y \in Y\bigr\}. \tag{1.2.6}$$

The formulation (1.2.4) clearly illustrates that the difficulties in solving a two-stage stochastic program with recourse lies in the recourse function $\mathcal{Q}$ — when this function is known, problem (1.2.4) is nothing but an ordinary non-linear programming problem.

With an absolutely continuous distribution of random parameters, though, even the evaluation of the integral in (1.2.5) may be cumbersome, and in fact most practical solution procedures for two-stage stochastic programs have to rely on theoretical results, justifying the assumption that the distribution of random parameters is discrete with finite support. With a finite distribution of random parameters, uncertainty is assumed to be appropriately described by a finite number of *scenarios*, each scenario $s \in \{1, \ldots, S\}$ corresponding to an outcome (q^s, h^s, T^s, W^s) of random parameters. Employing this assumption and denoting for $s \in \{1, \ldots, S\}$ the probability of scenario s to actually occur by p^s, the stochastic programming problem (1.2.3) reduces to an ordinary, large-scale (mixed-integer) linear programming problem,

$$
\begin{aligned}
\min\ & cx + \sum_{s=1}^{S} p^s q^s y^s, \\
\text{s.t. } & Ax \geq b, \\
& T^s x + W^s y^s \geq h^s, \quad && s = 1, \ldots, S, \\
& x \in X, y^s \in Y, && s = 1, \ldots, S.
\end{aligned}
$$

The size of this problem, however, usually prohibits the use of standard software for its solution, and hence specialized solution procedures based on decomposition techniques are required.

1.2.3 Multistage Stochastic Programs with Recourse

Obviously the decision process outlined in the previous section may be generalized to include multiple stages. In this section we consider a finite horizon decision process with an alternating sequence of decisions and observations of random data. For each stage $t = 1, \ldots, H$ we denote by x_t the vector of decisions to be made in stage t and by ξ_t the vector of random data that is observed in stage t. The multistage decision process may now be summarized as follows:

$$
\begin{array}{c}
\text{observation of } \xi_1 \\
\downarrow \\
\text{decision on } x_1 \\
\downarrow \\
\text{observation of } \xi_2 \\
\downarrow \\
\text{decision on } x_2 \\
\vdots \\
\downarrow \\
\text{observation of } \xi_H \\
\downarrow \\
\text{decision on } x_H
\end{array}
$$

For ease of notation we will assume throughout this section that the constraints of the problem have a Markovian structure, in the sense that decisions in stage t are directly coupled only to decisions in stage $t-1$ for $t = 2, \ldots, H$.

In the formulation of the two-stage recourse problem (1.2.3), non-anticipativity of the first-stage solution was implicitly represented in the model by the fact that the first-stage decision x is not allowed to depend on the random event ω. For the multistage model presented next, on the other hand, we choose to use explicit non-anticipativity constraints, stating that the decision in some stage t may only depend on the information available at this point in time. To formalize matters we need to enhance the notation somewhat. As mentioned above, for $t = 1, \ldots, H$, we let $\xi_t : \Omega \mapsto \mathbb{R}^{N_t}$ be a random vector defined on some probability space $(\Omega, \mathcal{F}, P)$, such that ξ_t is constituted by the random components of the stage t data. Furthermore, for $t = 1, \ldots, H$, we let $\xi_{[1,t]} = (\xi_1, \ldots, \xi_t)$ be the random vector of information available in stage t, and we denote by $\mathcal{F}_t \subseteq \mathcal{F}$ the σ-algebra generated by $\xi_{[1,t]}$ (see Appendix A). Assuming that data for the first stage is deterministic (i.e. ξ_1 is observed before the decision on x_1) we have $\mathcal{F}_1 = \{\emptyset, \Omega\}$, and obviously we have $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for $t = 1, \ldots, H-1$. Finally, for $t = 1, \ldots, H$, we denote by $x_t(\omega)$ the decision in stage t given the random event $\omega \in \Omega$. Now, for $t = 1, \ldots, H$, the non-anticipativity condition, requiring the decision in stage t to depend only on the information available in stage t, is equivalent to measurability of x_t with respect to $\mathcal{F}_t$. Thus the problem may be stated as

$$
\begin{aligned}
&\min \mathbb{E}\big[c_1(\omega)x_1(\omega) + \cdots + c_H(\omega)x_H(\omega)\big], && \\
&\text{ s.t. } W_1(\omega)x_1(\omega) \geq h_1(\omega), && P-a.s., \\
&\qquad T_t(\omega)x_{t-1}(\omega) + W_t(\omega)x_t(\omega) \geq h_t(\omega), && P-a.s.,\ t = 2, \ldots, H, \\
&\qquad x_t(\omega) \in X_t, && P-a.s.,\ t = 1, \ldots, H, \\
&\qquad x_t \text{ measurable with respect to } \mathcal{F}_t, && t = 1, \ldots, H.
\end{aligned}
$$

Here $c_t : \Omega \mapsto \mathbb{R}^{n_t}$, $h_t : \Omega \mapsto \mathbb{R}^{m_t}$, $T_t : \Omega \mapsto \mathbb{R}^{m_t \times n_{t-1}}$, and $W_t : \Omega \mapsto \mathbb{R}^{m_t \times n_t}$ are assumed to be measurable mappings defined on the probability space $(\Omega, \mathcal{F}, P)$, the components of which constitute the random vector ξ_t for $t = 1, \ldots, H$. As stated above we assume that c_1, h_1, and W_1 are deterministic, i.e. that $\mathcal{F}_1 = \{\emptyset, \Omega\}$, and hence the requirement that x_1 is measurable with respect to $\mathcal{F}_1$ means that x_1 must be constant for P-almost all $\omega \in \Omega$.

Again, the dynamics of the decision problem is clearly illustrated by an alternative dynamic programming formulation,

$$
\min\big\{c_1x_1 + \mathcal{Q}_2(x_1, \xi_1) \mid W_1x_1 \geq h_1,\ x_1 \in X_1\big\},
$$

where the expected recourse functions are given for $t = 2, \ldots, H-1$ by

$$
\begin{aligned}
\mathcal{Q}_t(x_{t-1}, \xi_{[1,t-1]}) = \mathbb{E}_{\xi_{[1,t]}|\xi_{[1,t-1]}}\big[&\min\big\{c_t(\omega)x_t + \mathcal{Q}_{t+1}(x_t, \xi_{[1,t]}(\omega)) \mid \\
&W_t(\omega)x_t \geq h_t(\omega) - T_t(\omega)x_{t-1},\ x_t \in X_t\big\}\big],
\end{aligned}
$$

and for the final stage by

$$
\begin{aligned}
\mathcal{Q}_H(x_{H-1}, \xi_{[1,H-1]}) = \mathbb{E}_{\xi_{[1,H]}|\xi_{[1,H-1]}}\big[&\min\big\{c_H(\omega)x_H \mid \\
&W_H(\omega)x_H \geq h_H(\omega) - T_H(\omega)x_{H-1},\ x_H \in X_H\big\}\big].
\end{aligned}
$$

Here, for $t = 2, \ldots, H$, we denote by $\mathbb{E}_{\xi_{[1,t]}|\xi_{[1,t-1]}}$ the (regular) conditional expectation with respect to the distribution of $\xi_{[1,t]}$, given $\xi_{[1,t-1]}$.

Just as for the two-stage problem, practical solution procedures for multistage stochastic programs generally rely on theoretical results justifying the assumption that the distribution of the random vector $\xi = \xi_{[1,H]}$ is discrete with finite support, say $\Xi = \{\bar{\xi}^1, \dots, \bar{\xi}^S\}$. We now employ this assumption and denote by $p^1, \dots, p^S$ the corresponding probabilities. For $t = 1, \dots, H$, the σ-algebra $\mathcal{F}_t$ then corresponds to a partition Π_t of Ξ such that each $\Xi_t \in \Pi_t$ represents a specific history $\xi_{[1,t]}$ of outcomes up to stage t, i.e. for all $\bar{\xi} \in \Xi_t$ we have $\bar{\xi}_{[1,t]} = \xi_{[1,t]}$. The probability of this history is,

$$p_{\xi_{[1,t]}} = \sum_{s:\bar{\xi}^s_{[1,t]} = \xi_{[1,t]}} p^s.$$

Furthermore, for $t = 1, \dots, H-1$ we define the set of *descendants* of the history $\xi_{[1,t]}$ as $\mathcal{D}_{\xi_{[1,t]}} = \left\{ \bar{\xi}_{[1,t+1]} \mid \bar{\xi} \in \Xi,\ \bar{\xi}_{[1,t]} = \xi_{[1,t]} \right\}$. Now, for $t = 2, \dots, H-1$, the expected recourse functions may be conveniently reformulated, writing the conditional expectation as a simple weighted sum,

$$\mathcal{Q}_t(x_{t-1}, \bar{\xi}_{[1,t-1]}) = \sum_{\bar{\xi}_{[1,t]} \in \mathcal{D}_{\xi_{[1,t-1]}}} \frac{p_{\bar{\xi}_{[1,t]}}}{p_{\xi_{[1,t-1]}}} \min\big\{ \bar{c}_t x_t + \mathcal{Q}_{t+1}(x_t, \bar{\xi}_{[1,t]}) \mid \\ \bar{W}_t x_t \geq \bar{h}_t - \bar{T}_t x_{t-1},\ x_t \in X_t \big\},$$

and correspondingly for the final stage,

$$\mathcal{Q}_H(x_{H-1}, \xi_{[1,H-1]}) = \sum_{\bar{\xi}_{[1,H]} \in \mathcal{D}_{\xi_{[1,H-1]}}} \frac{p_{\bar{\xi}_{[1,H]}}}{p_{\xi_{[1,H-1]}}} \min\big\{ \bar{c}_H x_H \mid \\ \bar{W}_H x_H \geq \bar{h}_H - \bar{T}_H x_{H-1},\ x_H \in X_H \big\}.$$

In this situation it is customary to think of uncertainty in terms of a *scenario tree*, the nodes of which are organized in H columns corresponding to the individual stages. The leaves of the tree (nodes in column H) correspond to the scenarios $\bar{\xi}^1, \dots, \bar{\xi}^S$, and for $t = 1, \dots, H-1$ the nodes in column t corresponds to elements of the partition Π_t, i.e. to some specific history of events up to time t. An example of a small scenario tree is illustrated in Figure 1.1.

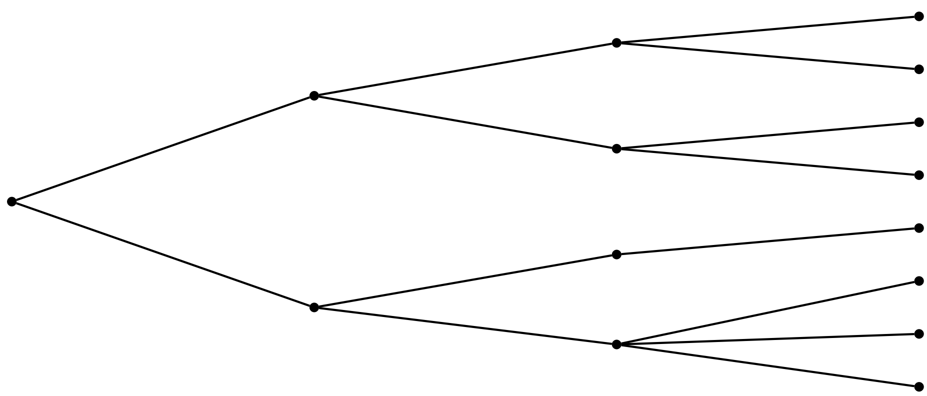

Figure 1.1: Scenario tree for a problem with 4 stages and 8 scenarios.

1.3 Summary of the Thesis

The remainder of this thesis is organized in three major parts. Chapters 2 and 3 form the first part which is concerned with classical stochastic programs with recourse. Hence we consider the general model that was presented in Section 1.2.2, and in particular the probability distribution of random parameters is assumed to be known and the expected value of total cost is minimized. This first part of the thesis is based on well-known results in stochastic programming, and hence the exposition is concise and proofs are omitted. In the second part of the thesis, comprised by Chapters 4 and 5, we shift attention toward two classes of non-classical stochastic programs with recourse. The models considered here are non-classical in the sense that, in the first case a more appropriate objective for minimization than the expectation-based one is found, and in the second case the probability distribution of random parameters is not assumed to be completely known. Finally, the third part of the thesis consists of Chapters 6 through 10 where five different applications of stochastic programming models in connection with capacity expansion of communication networks are presented. The second and third part contain the new contributions of this thesis to the field of stochastic programming.

Chapter 2 is concerned with the classical two-stage stochastic program with linear recourse, i.e. problem (1.2.4) with $Y = \mathbb{R}_+^{n_2}$. We discuss structural properties of the problem, primarily related to the expected recourse function (1.2.5). These properties include basic results on finiteness, continuity and convexity, but also more theoretical results, required to establish certain stability properties of the problem. These stability results take the form of continuity properties of the optimal-value function and the solution set mapping as functions of the underlying probability distribution. Such results may be employed, in particular, to justify the assumption that the probability distribution is discrete with finite support. This assumption is fundamental for the decomposition-based solution procedures discussed in conclusion of the chapter.

Next, in Chapter 3 we discuss the classical two-stage stochastic program with mixed-integer recourse, i.e. problem (1.2.4) with $Y = \mathbb{R}_+^{\bar{n}_2} \times \mathbb{Z}_+^{\tilde{n}_2}$, where $\bar{n}_2 + \tilde{n}_2 = n_2$. Again the focus is on theoretical properties concerning finiteness and continuity of the expected recourse function and stability of optimal solutions, ultimately leading to the presentation of a number of alternative decomposition-based solution procedures.

In Chapter 4, which is based on Riis and Schultz [116], we discuss a non-classical two-stage stochastic program with linear recourse, referred to as the minimum risk problem. The problem is formulated in an attempt to give a more adequate description of risk aversion than what is provided by the classical stochastic recourse problem. It consists in minimizing the probability of total cost exceeding a certain prescribed threshold value, that may be thought of as e.g. a budget limit or the level of bankruptcy. It is easily seen that the problem is in fact equivalent to a classical two-stage stochastic program with mixed-integer recourse, but we show that a number of structural properties may be established with less restrictive assumptions than what is usually employed for the analysis of the classical problem. In line with the presentation in Chapters 2 and 3 we present results in the form of continuity properties of the recourse function and stability properties of optimal solutions. Also, a specialized solution procedure is elaborated and results of our computational experiments are reported.

Another non-classical stochastic recourse program, referred to as the minimax approach to stochastic programming, is considered in Chapter 5 which is based on Riis and Andersen [113]. Based on the fact that the probability distribution of random parameters will hardly ever be directly accessible, this approach rests on the establishment of a set of possible or conceivable probability distributions, that are consistent with the available information. A recourse function, that may be defined for example according to either of the models discussed in Chapters 2, 3, or 4, is then minimized with respect to the worst of these conceivable distributions. In contrast to the minimum risk problem discussed in Chapter 4, the minimax problem has been the subject of a significant amount of research during a number of years, but only a few practical solution procedures have been proposed. In Chapter 5, we show how structural properties established for the problems considered in Chapters 2, 3, and 4 may be employed to arrive at stability results for the corresponding minimax problem. These stability results justify, in particular, a simplifying assumption, that allows us to restrict attention to discrete probability measures. Based on this assumption, we show how well-known algorithms for the different problem classes may be extended to solve the corresponding minimax problem.

With Chapter 6, which is based on Riis and Andersen [112], we start our presentation of applications of stochastic programming in telecommunications. The problem that we consider in Chapter 6 is the simplest one in this respect, concerning the multiperiod capacity expansion of a single telecommunications connection. Our starting point is a two-stage formulation of the problem presented by Laguna [81], who also proposed a solution procedure. In Chapter 6, we present a new preprocessing rule that drastically reduces computation time for this algorithm. Also, we discuss the alternative of using a multistage formulation of the problem, and we elaborate a corresponding recursive solution procedure. Computational results are reported to document the effect of the preprocessing rule and to show the practicability of the multistage procedure.

In Chapter 7, which is based on Riis and Andersen [114], we consider the so-called capacitated network design problem, which has been studied extensively in a deterministic setting. The problem is to install capacity on links of a telecommunications network in modularities of two fixed batch sizes. The capacity expansion must be carried out so as to meet customer demand while minimizing total costs incurred. We propose a two-stage stochastic programming formulation with integer first stage and continuous second stage, and discuss how valid inequalities derived for the deterministic problem may be generalized for the stochastic program. Also, a branch-and-cut algorithm, employing the valid inequalities as cutting planes, is elaborated, and we report results of a series of computational experiments performed on two real-life instances.

Chapter 8 is based on Riis and Lodahl [115]. It concerns the capacity expansion of a telecommunications network in the face of uncertain future demand and potential future failures of network components. The problem is formulated as a bicriterion stochastic program with recourse in which the total cost of the capacity expansion and the probability of future capacity requirements to be violated are simultaneously minimized. Here the second objective is a special case of the one considered in Chapter 4. We elaborate a solution procedure that determines all non-dominated solutions to the problem by a reduced feasible region method. During the course of this procedure a sequence of subproblems are solved by an algorithm, that is in many ways similar to the one presented in

Chapter 4. Computational results are reported for three real-life instances, one of which is a problem faced by SONOFON, a Danish communications network operator.

In Chapter 9, which is based on Riis, Skriver, and Lodahl [117], we consider a network design problem arising in mobile communications. At the core of the network is a number of mobile switching centers (MSCs), each serving a number of base station controllers (BSCs). The network design problem involves the deployment of a number of new MSCs, the allocation of BSCs to new and existing MSCs, and the capacity expansion of transmission links interconnecting the MSCs. We formulate the problem as a two-stage stochastic program with mixed-integer recourse. To solve the problem we apply a decomposition procedure, solving scenario subproblems by means of branch-and-cut. The solution procedure has been tested on a real-life problem instance provided by SONOFON, and we report results of our computational experiments.

Finally, Chapter 10 is based on Riis, Skriver, and Møller [118]. This is a case study concerning the design and dimensioning of the IP (internet protocol) network of TDC, the largest Danish communications network operator. Due to historical reasons, the number of IP POPs (points of presence) in the network has reached a level believed to be too high. To point out potential IP POPs for dismantling, we consider a network planning problem concerning dimensioning of the IP POPs, connection of customers to the network, and capacity expansion of transmission links interconnecting the IP POPs. The problem is formulated as a two-stage stochastic program with linear recourse, and a cutting plane procedure is elaborated to solve it. Computational results are reported for the IP network of TDC.

Part I

Classical Stochastic Programs

Chapter 2

Two-Stage Stochastic Programs with Linear Recourse

The two-stage stochastic program with linear recourse is probably the single most prevalent problem within the class of stochastic programming models. In this chapter we give an account of the most important theoretical properties of the problem, and in particular we discuss the line of research related to stability of optimal solutions when the underlying probability distribution is subjected to perturbations. Also, we present the well-known L-shaped algorithm for the problem and discuss a few of its extensions. For an overview of the research in the field of stochastic linear programming, we refer to the textbooks by Birge and Louveaux [22], Kall and Wallace [67] and Prékopa [107]. See also the extensive online bibliography of van der Vlerk [157].

2.1 Problem Formulation

The classical two-stage stochastic program with recourse was introduced in Section 1.2.2. We recall that the problem consists in the determination of a first-stage decision that must be made without certain knowledge on the random parameters of the model, and so as to minimize the sum of direct cost and the expected value of future cost. The future cost is determined through a second-stage problem, as certain recourse actions can be taken after uncertainty has been revealed. In this chapter we study the problem under the assumption that both the first-stage and the second-stage problem may be modeled appropriately as linear programming problems. Moreover, because very few theoretical results are available for problems with a random recourse matrix, we will assume that the recourse matrix is fixed — we say that the problem has fixed recourse. We refer to Walkup and Wets [160] for a discussion of problems with a random recourse matrix and the difficulties this may give rise to.

Since we are going to address the issue of stability of the stochastic program when the underlying probability distribution is subjected to perturbations, we will be interested in continuity properties of the expected recourse function as a function of the probability distribution as well as of the first-stage decision. To facilitate such an analysis, it will be convenient to formulate the model slightly different from what we did in Section 1.2.2. We let $\xi : \Omega \mapsto \mathbb{R}^N$ be a random vector defined on some probability space $(\Omega, \mathcal{F}, P)$. The components of ξ constitute the random second-stage data, consisting of the second-stage cost $\tilde{q} : \Omega \mapsto \mathbb{R}^{n_2}$, the second-stage right-hand side $\tilde{h} : \Omega \mapsto \mathbb{R}^{m_2}$, and the technology matrix $\tilde{T} : \Omega \mapsto \mathbb{R}^{m_2 \times n_1}$. More precisely we have $N = n_2 + m_2(1 + n_1)$, and for $\omega \in \Omega$ we have $\xi(\omega) = \big(\tilde{q}(\omega), \tilde{h}(\omega), \tilde{T}_1(\omega), \ldots, \tilde{T}_{m_2}(\omega)\big)$ where $\tilde{T}_k$ denotes the kth row of $\tilde{T}$ for $k = 1, \ldots, m_2$. Introducing the induced Borel probability measure $\mu = P \circ \xi^{-1}$ on $\mathbb{R}^N$, the two-stage linear recourse problem may now be given the following formulation,

$$TSLR(\mu) \qquad \min\big\{cx + \mathcal{Q}(x, \mu) \mid Ax = b,\ x \in \mathbb{R}^{n_1}_+\big\}, \tag{2.1.1}$$

where the expected recourse function $\mathcal{Q}$ explicitly depends on the distribution μ,

$$\mathcal{Q}(x, \mu) = \mathbb{E}[\Phi(x, \xi)] = \int_{\mathbb{R}^N} \Phi(x, \xi)\mu(d\xi), \tag{2.1.2}$$

and the second-stage value function Φ is given by

$$\Phi(x, \xi) = \min\big\{q(\xi)y \mid Wy = h(\xi) - T(\xi)x,\ y \in \mathbb{R}^{n_2}_+\big\}. \tag{2.1.3}$$

Here $c \in \mathbb{R}^{n_1}$ and $b \in \mathbb{R}^{m_1}$ are known vectors and $A \in \mathbb{R}^{m_1 \times n_1}$ and $W \in \mathbb{R}^{m_2 \times n_2}$ are known matrices. The second-stage cost $q : \mathbb{R}^N \mapsto \mathbb{R}^{n_2}$, the second-stage right-hand side $h : \mathbb{R}^N \mapsto \mathbb{R}^{m_2}$, and the technology matrix $T : \mathbb{R}^N \mapsto \mathbb{R}^{m_2 \times n_1}$, on the other hand, are represented as mappings picking out the appropriate components of the random vector ξ.

Remark 2.1.1. In the literature one often sees q, h, and T defined as affine mappings, i.e.

$$\begin{aligned} q(\xi) &= \bar{q}^0 + \sum_{i=1}^N \bar{q}^i \xi_i, \\ h(\xi) &= \bar{h}^0 + \sum_{i=1}^N \bar{h}^i \xi_i, \\ T(\xi) &= \bar{T}^0 + \sum_{i=1}^N \bar{T}^i \xi_i, \end{aligned}$$

where $\bar{q}^i$, $\bar{h}^i$, and $\bar{T}^i$ are known vectors and matrices, respectively, for $i = 0, \ldots, N$. For our purposes, however, it will be more convenient to think simply of ξ as constituted by the components of q, h, and T as explained above, so that μ is the joint distribution of the second-stage cost, the second-stage right-hand side, and the technology matrix.

Remark 2.1.2. An important special case of the stochastic program (2.1.1)-(2.1.3) is the situation when $W = [I, -I]$ where I is the $m_2 \times m_2$-identity matrix. This case is referred to as simple recourse. Stochastic programs with simple recourse have received particular attention in the literature, since the special structure of the problem provides significant computational advantages. We refer to research papers by e.g. Wets [164] and Ziemba [168], and to the PhD-thesis by van der Vlerk [156] for results in this direction.

2.2 Structural Properties

In this section we discuss the definition and structure of certain feasibility sets, we present structural properties of the expected recourse function, and we consider the stability of optimal solutions when the underlying probability measure is subjected to perturbations.

2.2.1 Feasibility Sets

For now, we consider the stochastic program (2.1.1) for some known and fixed Borel probability measure $\mu \in \mathcal{P}(\mathbb{R}^N)$, and we denote by $\Xi \subseteq \mathbb{R}^N$ the support of μ, i.e. the smallest closed subset such that $\mu(\Xi) = 1$. We denote the set of first-stage solutions that satisfy the deterministic constraints of (2.1.1) by K_1. That is, we let

$$K_1 = \{x \in \mathbb{R}^{n_1}_+ \mid Ax = b\}.$$

Now, for all practical purposes, it does not make sense to consider those first-stage solutions in K_1 that lead to an infeasible second-stage problem (2.1.3) with a positive probability, since in that case the expected recourse function (2.1.2) is infinite. Therefore, the first-stage solutions should satisfy not only the deterministic constraints, $x \in K_1$, but also a set of *induced constraints*, $x \in K_2$, where we define

$$K_2 = \{x \in \mathbb{R}^{n_1}_+ \mid \mathcal{Q}(x, \mu) < +\infty\}.$$

This definition, however, is obviously not particularly useful when it comes to characterizing feasible first-stage solutions, and in particular it does not provide a means of checking whether a given first-stage solution $x \in K_1$ satisfies the induced constraints without having to actually compute $Q(x, \mu)$. To this end, it is convenient to introduce the positive cone generated by W,

$$\operatorname{pos} W = \{t \in \mathbb{R}^{m_2} \mid \exists\, y \in \mathbb{R}^{n_2}_+ : Wy = t\},$$

and consider the following alternative feasibility set,

$$K_2^P = \{x \in \mathbb{R}^{n_1}_+ \mid \forall \xi \in \Xi : h(\xi) - T(\xi)x \in \operatorname{pos} W\}.$$

The set K_2^P possesses some appealing theoretical properties.

Theorem 2.2.1. *The set K_2^P is*

(a) closed and convex;

(b) polyhedral if the support Ξ of μ is a finite set.

It is easily seen that we always have $K_2 \subseteq K_2^P$. For the opposite inclusion, however, additional assumptions are required. In particular we need the following definition.

Definition 2.2.1. If μ is such that for all $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$ and $k = 1, \ldots, m_2$, the first moments of $q_j(\xi)h_k(\xi)$ and $q_j(\xi)T_{ki}(\xi)$ are finite, then μ is said to satisfy the weak covariance condition.

The following result may be found in Wets [163].

Theorem 2.2.2. *If μ satisfies the weak covariance condition, then we have $K_2 = K_2^P$.*

The weak covariance condition is easily seen to hold in a number of specific cases, for example when μ has finite second moments, when q is fixed and μ has finite first moments, or when (h, T) is fixed and μ has finite first moments. Thus, in any of these cases, the alternative formulation K_2^P may provide a convenient way of checking whether any induced constraints are violated by some first-stage solution $x \in K_1$.

In practice the problem concerning induced constraints is often simplified because the second-stage problem is feasible for all possible right-hand sides. Even when this is not the case, it may be that a feasible second-stage solution exists for all right-hand sides that may actually occur (or at least with probability one), i.e. that the second-stage problem is feasible for all first-stage solutions $x \in K_1$ and all outcomes of the random vector $\xi \in \Xi$. These situations are referred to as *complete recourse* and *relatively complete recourse*, respectively.

Definition 2.2.2. A stochastic program with fixed linear recourse is said to have

(i) complete recourse if $\operatorname{pos} W = \mathbb{R}^{m_2}$;

(ii) relatively complete recourse if $h(\xi) - T(\xi)x \in \operatorname{pos} W$ for all $x \in K_1$ and all $\xi \in \Xi$.

Note that with complete recourse we have $K_2^P = \mathbb{R}^{n_1}_+$, whereas with relatively complete recourse we have $K_1 \subseteq K_2^P$. Hence we see that whenever μ satisfies the weak covariance condition, either complete recourse or relatively complete recourse is sufficient to ensure $K_1 \subseteq K_2$ and hence in these cases there are no induced constraints.

2.2.2 The Expected Recourse Function

In the following we will consider the stochastic program (2.1.1) for probability measures $\mu \in \mathcal{P}(\Xi)$, where $\Xi \subseteq \mathbb{R}^N$ is some closed set. We will employ the following assumptions.

(A1) For all $t \in \mathbb{R}^{m_2}$ there exists $y \in \mathbb{R}^{n_2}_+$ such that $Wy = t$.

(A2) For all $\xi \in \Xi$ there exists $u \in \mathbb{R}^{m_2}$ such that $uW \leq q(\xi)$.

Remark 2.2.1. Note that assumption (A2) is the reason for us to restrict attention to probability measures with support within some closed set $\Xi \subseteq \mathbb{R}^N$. In particular, it is not unreasonable to assume that all possible outcomes of the random second-stage cost meets (A2), since the model would probably otherwise be incorrectly specified — practical problems are usually not unbounded. It would, however, certainly be an unreasonably strong assumption to require (A2) to be fulfilled for all $\xi \in \mathbb{R}^N$.

(A1) is the assumption of complete recourse, discussed in the previous section, whereas (A2) is the assumption of dual feasibility of the second-stage problem for all possible outcomes $\xi \in \Xi$. These two assumptions ensure feasibility and boundedness, respectively, of the second-stage problem, and hence they are sufficient to establish the following properties of the second-stage value-function defined by (2.1.3).

Lemma 2.2.1. *Assume (A1)-(A2). Then Φ is a real-valued function on $\mathbb{R}^{n_1} \times \Xi$ and*

(a) a convex, piecewise linear function of x for all $\xi \in \Xi$;

(b) a convex, piecewise linear function of $(h(\xi), T(\xi))$ for all $x \in \mathbb{R}^{n_1}$.

(c) a concave, piecewise linear function of $q(\xi)$ for all $x \in \mathbb{R}^{n_1}$.

If, in addition to assumptions (A1) and (A2), the probability measure satisfies the weak covariance condition, then the expected recourse function is finite on $\mathbb{R}^{n_1}$, and since the expectation-operator is linear, the expected recourse function inherits its structural properties from the second-stage value function. Hence we have the following.

Theorem 2.2.3. *Assume (A1)-(A2). If $\mu \in \mathcal{P}(\Xi)$ satisfies the weak covariance condition, then $\mathcal{Q}(\cdot, \mu)$ is a real-valued, Lipschitzian, and convex function on $\mathbb{R}^{n_1}$. Moreover,*

(a) if the support of μ is a finite set, then $\mathcal{Q}(\cdot, \mu)$ is piecewise linear on $\mathbb{R}^{n_1}$;

(b) if μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^N$, then $\mathcal{Q}(\cdot, \mu)$ is differentiable on $\mathbb{R}^{n_1}$.

Remark 2.2.2. Clearly, the assumption of complete recourse in Theorem 2.2.3 may be replaced by the weaker assumption of relatively complete recourse or may even be removed, in which case, however, the results should be stated for $x \in K_2$ only.

As previously mentioned, in order to arrive at stability results for the stochastic program (2.1.1), we will need results regarding the joint continuity of the expected recourse function with respect to the first-stage decision and the probability distribution. To this end, we recall that the set of Borel probability measures is endowed with the notion of weak convergence (see Appendix A). Now, since $\mathcal{Q}(\cdot, \mu)$ is convex for $\mu \in \mathcal{P}(\Xi)$, joint continuity of $\mathcal{Q}$ with respect to x and μ is implied by continuity of $\mathcal{Q}(x, \cdot)$ for all $x \in X$. (See e.g. Rockafellar [123, Theorem 10.7].) The continuity of $\mathcal{Q}$ with respect to μ, on the other hand, requires some uniform integrability condition to be satisfied cf. e.g. Billingsley [18, Theorem 5.4] or Hoffman-Jørgensen [61, Section 5.2]). In particular, Robinson and Wets [122] considered the following extension of uniform integrability, saying that a family $\mathcal{F}$ of real-valued continuous functions on Ξ is uniformly integrable with respect to some family $\mathcal{P} \subseteq \mathcal{P}(\Xi)$ of Borel probability measures, if for all $\epsilon > 0$ there exists a compact set $A \subseteq \Xi$ such that for all $f \in \mathcal{F}$ and $\mu \in \mathcal{P}$ we have

$$\int_{\Xi \setminus A} |f(\xi)| \mu(d\xi) < \epsilon.$$

Now, assuming uniform integrability of the family of functions $\mathcal{F} = \{\Phi(x, \cdot) \mid x \in \mathbb{R}^{n_1}\}$ with respect to some family $\mathcal{P}$ of Borel probability measures, Robinson and Wets proved joint continuity of $\mathcal{Q}$ on $\mathbb{R}^{n_1} \times \mathcal{P}$.

As pointed out by Römisch and Schultz [125], the above uniform integrability of the family of recourse integrands $\{\Phi(x, \cdot) \mid x \in \mathbb{R}^{n_1}\}$ with respect to some family $\mathcal{P}$ of Borel probability measures is achieved, for example if $\mathcal{P}$ is defined to be the following family of Borel probability measures,

$$\mathcal{P}_{p,K}(\Xi) = \left\{ \mu \in \mathcal{P}(\Xi) \;\middle|\; \int_{\Xi} ||\xi||^p \mu(d\xi) \leq K \right\},$$

for some real numbers $p > 2$ and $K > 0$.

Theorem 2.2.4. *Assume (A1)-(A2) and let $\mu \in \mathcal{P}_{p,K}(\Xi)$ for some $p > 2$ and $K > 0$. Then $\mathcal{Q}$, as a function from $\mathbb{R}^{n_1} \times \mathcal{P}_{p,K}(\Xi)$ to $\mathbb{R}$, is continuous on $\mathbb{R}^{n_1} \times \{\mu\}$.*

Remark 2.2.3. As mentioned above, Römisch and Schultz [125] pointed out that the integrability assumption in Theorem 2.2.4 is, in general, more restrictive than what is required here. We find this assumption convenient, though, because it allows us to supplement the qualitative continuity result of Theorem 2.2.4 with a quantitative continuity result cf. Theorem 2.2.5 below, and because it readily extends to the case of mixed-integer recourse cf. Theorem 3.2.4 on page 32. For examples of weaker integrability assumptions leading to the joint continuity of $\mathcal{Q}$, we refer to e.g. Kall [65] and Robinson and Wets [122].

Theorem 2.2.4 leads directly to the qualitative stability results presented in the following section. We will, however, also be interested in quantitative stability results for the stochastic program (2.1.1). These results rely on the identification of a suitable metric on $\mathcal{P}(\Xi)$, that preferably metrizes (at least locally) weak convergence. One such metric is the bounded Lipschitz metric β which was considered in Römisch and Schultz [125]. This metric is defined for $\mu, \nu \in \mathcal{P}(\Xi)$ by

$$\beta(\mu,\nu) = \sup_{g:\Xi\mapsto\mathbb{R}} \left\{ \int_\Xi g(\xi)\mu(d\xi) - \int_\Xi g(\xi)\nu(d\xi) \;\middle|\; \sup_{\xi\in\Xi}|g(\xi)| + \sup_{\substack{\xi,\tilde{\xi}\in\Xi\\ \xi\neq\tilde{\xi}}} \frac{|g(\xi)-g(\tilde{\xi})|}{||\xi-\tilde{\xi}||} \leq 1 \right\}.$$

For details about the bounded Lipschitz metric we refer to Dudley [42] or Rachev [108]. We now have the following Hölder estimate for the expected recourse function.

Theorem 2.2.5. *Assume (A1)-(A2), let* $D \subseteq \mathbb{R}^{n_1}$ *be non-empty and compact, and let* $\mu \in \mathcal{P}_{p,K}(\Xi)$ *for some* $p > 2$ *and* $K > 0$*. Then there exists* $L > 0$ *such that*

$$\big|\mathcal{Q}(x,\mu) - \mathcal{Q}(z,\nu)\big| \leq L \cdot \big(||x-z|| + \beta(\mu,\nu)^{\frac{p-1}{p}}\big)$$

for all $x, z \in D$ *and all* $\nu \in \mathcal{P}_{p,K}(\Xi)$.

Remark 2.2.4. Since the bounded Lipschitz metric is known to metrize weak convergence on $\mathcal{P}(\Xi)$, Theorem 2.2.5 quantifies the result in Theorem 2.2.4.

As already mentioned, alternative quantitative results related to different probability metrics are available. For example, Römisch and Schultz [126, 127] considered the so-called L_p-Wasserstein metric W_p which is known to majorize the bounded Lipschitz metric. The L_p-Wasserstein metric is defined for $\mu, \nu \in \mathcal{M}_p(\Xi)$ by

$$W_p(\mu,\nu) = \left(\inf_{\eta\in\mathcal{P}(\Xi\times\Xi)} \left\{ \int_{\Xi\times\Xi} ||\xi-\tilde{\xi}||^p \eta(d\xi,d\tilde{\xi}) \;\middle|\; \eta\circ\pi_1^{-1} = \mu,\ \eta\circ\pi_2^{-1} = \nu \right\} \right)^{\frac{1}{p}}$$

where π_1 and π_2 are the first and second projections, respectively, and

$$\mathcal{M}_p(\Xi) = \left\{ \mu \in \mathcal{P}(\Xi) \;\middle|\; \int_\Xi ||\xi||^p \mu(d\xi) < \infty \right\}.$$

For this metric it holds that $(\mathcal{M}_p(\Xi), W_p)$ is a metric space and if $\mu \in \mathcal{M}_p(\Xi)$ and $\mu_n \in \mathcal{M}_p(\Xi)$ for $n \in \mathbb{N}$, then we have $W_p(\mu,\mu_n) \xrightarrow{n\to\infty} 0$ if and only if $\mu \xrightarrow{w} \mu$ and $\int_\Xi ||\xi||^p \mu_n(d\xi) \xrightarrow{n\to\infty} \int_\Xi ||\xi||^p \mu(d\xi)$. (See e.g. Givens and Shortt [50] or Rachev [108] for more details about the metric.) With the L_2 -Wasserstein metric, we have the following Lipschitz estimate for the expected recourse function.

Theorem 2.2.6. *Assume (A1)-(A2), let $D \subseteq \mathbb{R}^{n_1}$ be non-empty and compact, and let $\mu \in \mathcal{M}_2(\Xi)$. Then there exist $L > 0$ and $\delta > 0$ such that*

$$\big|\mathcal{Q}(x,\mu) - \mathcal{Q}(z,\nu)\big| \leq L \cdot \big(||x - z|| + W_2(\mu,\nu)\big)$$

for all $x, z \in D$ and all $\nu \in \mathcal{M}_2(\Xi)$ with $W_2(\mu,\nu) < \delta$.

Remark 2.2.5. If the second-stage cost q is deterministic, the continuity results presented here may all be strengthened, in the sense that Theorem 2.2.4 and Theorem 2.2.5 can be stated for $p > 1$ rather than for $p > 2$, and likewise, Theorem 2.2.6 can be stated for the L_1-Wasserstein metric with $\mu, \nu \in \mathcal{M}_1(\Xi)$ rather than for the L_2-Wasserstein metric with $\mu, \nu \in \mathcal{M}_2(\Xi)$.

2.2.3 Stability

In many practical applications of stochastic programming the probability distribution of random parameters is not completely known, and hence the true distribution may have to be replaced in the model by some suitable estimate, such as e.g. empirical measures. Also, even if the true distribution μ of random parameters is known, the approximation of μ by simpler probability measures may be required to facilitate practical computations. In fact, as we will see in Section 2.3, most solution procedures for two-stage stochastic programs with linear recourse rely on the assumption that the probability distribution of random parameters is discrete with finite support, which will obviously often not be the case in practice. For these reasons, the issue of stability of the stochastic program (2.1.1), when the underlying probability distribution is subjected to perturbations, is an important one. For a collection of results on stability of stochastic programs with linear recourse along lines similar to the presentation here, we refer to e.g. Dupačová [45, 46], Kall [65], Robinson and Wets [122], Römisch and Schultz [125, 126, 127, 128], Römisch and Wakolbinger [130], Shapiro [146], and Wang [162]. We also note that our approach to the issue of stability builds on results obtained by Klatte [72, 73] in the analysis of parametric optimization problems. Finally, let us mention that Schultz [142] provides an excellent overview of results on stability in stochastic programming, covering also problems with mixed-integer recourse.

The stability results presented here take the form of continuity properties of the optimal-value function, $\varphi : \mathcal{P}(\Xi) \mapsto \mathbb{R}$, defined by

$$\varphi(\mu) = \inf\big\{cx + \mathcal{Q}(x,\mu) \mid Ax = b,\ x \in \mathbb{R}^{n_1}_+\big\},$$

and of the solution set mapping, $\Psi : \mathcal{P}(\Xi) \mapsto \mathbb{R}^{n_1}$, defined by

$$\Psi(\mu) = \arg\min\big\{cx + \mathcal{Q}(x,\mu) \mid Ax = b,\ x \in \mathbb{R}^{n_1}_+\big\}.$$

The qualitative continuity results stated in Theorem 2.2.4 leads directly to a qualitative stability result. More precisely, having established the joint continuity of $\mathcal{Q}$ with respect to x and μ it is straightforward to follow the lines of Berge [14] to prove continuity of φ and Berge upper semicontinuity of Ψ, cf. also the results of Bank et al. [9]. (Recall that the point-to-set mapping, Ψ, is Berge upper semicontinuous at some $\mu \in \mathcal{P}(\Xi)$ if for any open set $G \subseteq \mathbb{R}^{n_1}$ with $\Psi(\mu) \subseteq G$ there exists some neighborhood U of μ in $\mathcal{P}(\Xi)$ such that $\Psi(\nu) \subseteq G$ for all $\nu \in U$.) Hence we have the following.

Theorem 2.2.7. *Assume (A1)-(A2) and for some $p > 2$ and $K > 0$ let $\mu \in \mathcal{P}_{p,K}(\Xi)$ be such that $\Psi(\mu)$ is non-empty and bounded. Then,*

(a) φ as a function from $\mathcal{P}_{p,K}(\Xi)$ to $\mathbb{R}$ is continuous at μ;

(b) Ψ as a mapping from $\mathcal{P}_{p,K}(\Xi)$ to $\mathbb{R}^{n_1}$ is Berge upper semicontinuous at μ;

(c) there exists some neighborhood U of μ in $\mathcal{P}_{p,K}(\Xi)$ such that $\Psi(\nu)$ is non-empty for all $\nu \in U$.

Remark 2.2.6. As pointed out also in Remark 2.2.3, the integrability assumption in Theorem 2.2.7, restricting the probability measures to the set $\mathcal{P}_{p,K}(\Xi)$ for some $p > 2$ and $K > 0$, is more restrictive than what is generally required here. Again, all that is needed is some uniform integrability condition leading to the joint continuity of $\mathcal{Q}$ with respect to x and μ, cf. the results of e.g. Kall [65] and Robinson and Wets [122].

As previously pointed out, the true distribution of random parameters will not be completely known in many practical applications of stochastic programming, and hence the true distribution may have to be replaced by some suitable estimate. Let us consider for a moment the situation when the true distribution μ is approximated by empirical measures. In particular, we let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of independent and identically distributed N-dimensional random vectors defined on some probability space $(\Omega, \mathcal{F}, P)$, and we denote by $\mu \in \mathcal{P}(\Xi)$ their common distribution. This gives rise to a corresponding sequence of empirical probability measures on Ξ defined by

$$\mu_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\xi_i(\omega)},$$

where $\delta_{\xi_i(\omega)}$ denotes the measure with unit mass at $\xi_i(\omega)$ for $i = 1, \ldots, n$. It is well-known that we have $\mu_n(\omega) \xrightarrow{w} \mu$ for P-almost all $\omega \in \Omega$ cf. e.g. Dudley [42, Theorem 11.4.1]. Using this fact, it is now straightforward to apply Theorem 2.2.7 to obtain the following result.

Theorem 2.2.8. *Assume (A1)-(A2) and for some $p > 2$ and $K > 0$ let $\mu \in \mathcal{P}_{p,K}(\Xi)$ be such that $\Psi(\mu)$ is non-empty and bounded. Then,*

(a) $\varphi(\mu_n(\omega)) \xrightarrow{n \to \infty} \varphi(\mu)$ for P-almost all $\omega \in \Omega$;

(b) for any open set $G \subseteq \mathbb{R}^{n_1}$ with $\Psi(\mu) \subseteq G$ and for P-almost all $\omega \in \Omega$ there exists some $n_0(\omega) \in \mathbb{N}$ such that $\Psi(\mu_n(\omega)) \subseteq G$ for all $n \geq n_0(\omega)$;

(c) for P-almost all $\omega \in \Omega$ there exists some $n_1(\omega) \in \mathbb{N}$ such that $\Psi(\mu_n(\omega))$ is non-empty for all $n \geq n_1(\omega)$.

Finally, the quantitative continuity results stated in Theorem 2.2.5 and Theorem 2.2.6, respectively, may be applied to obtain the following quantitative stability results.

Theorem 2.2.9. *Assume (A1)-(A2) and for some $p > 2$ and $K > 0$ let $\mu \in \mathcal{P}_{p,K}(\Xi)$ be such that $\Psi(\mu)$ is non-empty and bounded. Then there exist $L > 0$ and $\delta > 0$ such that*

$$\big|\varphi(\mu) - \varphi(\nu)\big| \leq L \cdot \beta(\mu, \nu)^{\frac{p-1}{p}}$$

for all $\nu \in \mathcal{P}_{p,K}(\Xi)$ with $\beta(\mu, \nu) < \delta$.

Theorem 2.2.10. *Assume (A1)-(A2) and let $\mu \in \mathcal{M}_2(\Xi)$ be such that $\Psi(\mu)$ is non-empty and bounded. Then there exist $L > 0$ and $\delta > 0$ such that*

$$\big|\varphi(\mu) - \varphi(\nu)\big| \leq L \cdot W_2(\mu, \nu)$$

for all $\nu \in \mathcal{M}_2(\Xi)$ with $W_2(\mu, \nu) < \delta$.

Remark 2.2.7. In accordance with Remark 2.2.5 we note that if the second-stage cost q is deterministic, the stability results presented here may be strengthened, in the sense that Theorem 2.2.7, Theorem 2.2.8, and Theorem 2.2.9 can be stated for $p > 1$, and Theorem 2.2.10 can be stated for the L_1-Wasserstein metric with $\mu, \nu \in \mathcal{M}_1(\Xi)$.

Remark 2.2.8. Shapiro [146] quantified the result in Theorem 2.2.7 (b) for problems with fixed second-stage cost q and fixed technology matrix T. In this paper a second-order growth condition on the objective is employed to prove Lipschitz upper semicontinuity of Ψ. This result was extended by Römisch and Schultz [128] who provided a Lipschitz estimate for the Hausdorff distance of solution sets, assuming strong convexity of the objective. Furthermore, the authors discuss application of the results when the true distribution is estimated by empirical measures, cf. the discussion above.

2.3 Solution Procedures

A large majority of the specialized solution procedures that have been proposed in the literature for two-stage stochastic linear programs, rely on decomposition techniques, exploiting the special structure of the problem to break it up into smaller and more manageable pieces. For a recent comprehensive overview of decomposition methods in stochastic linear programming, we refer to Ruszczyński [133]. In general, a distinction is made between *primal decomposition* in which the problem is decomposed with respect to the individual stages, and *dual decomposition* in which the problem is decomposed with respect to the individual outcomes of random parameters. In this thesis we will primarily be concerned with primal decomposition procedures when it comes to the solution of two-stage stochastic programs with linear recourse. These solution procedures build on the so-called L-shaped algorithm, introduced in 1969 by Van Slyke and Wets [159] and discussed in detail in Section 2.3.1 below. This algorithm is based on Benders decomposition method (see Benders [12]) and forms an outer linerization of the problem, exploiting the fact that with a finite distribution of random parameters the recourse function is convex and piecewise linear on a polyhedral domain. We note that this approach is in fact equivalent to performing a Dantzig-Wolfe decomposition (see Dantzig and Wolfe [39]) of the dual of the problem, and hence the inner linerization method suggested by Dantzig and Madansky [38] in 1961 may be seen as a dual method to the L-shaped algorithm. Some of the most efficient solution procedures for two-stage stochastic programs as of today, may be seen as extensions of the L-shaped algorithm. Here we mention in particular the multicut version of the L-shaped algorithm proposed by Birge and Louveaux [21], the regularized decomposition procedure introduced by Ruszczyński [132] (see also Ruszczyński and Swietanowski [134]) and the stochastic decomposition procedure elaborated by Higle and Sen [57] (see also Higle and Sen [56, 58, 59, 60]). As for dual decomposition procedures for stochastic linear programs, the most prominent example is probably the progressive hedging algorithm introduced by Rockafellar and Wets [124], which is known as an operator splitting method. Other examples include methods based on an augmented Lagrangian decomposition as in e.g. Mulvey and Ruszczyński [95] and Rosa and Ruszczyński [131]. A particular advantage of these dual decomposition methods is that they have all been successfully applied to multistage stochastic linear programs. For completeness we should note that also a primal decomposition approach for multistage problems, generalizing the L-shaped method, has been proposed by Birge [19] and Pereira and Pinto [106].

2.3.1 L-shaped Decomposition

In this section we consider the L-shaped algorithm in some detail, since this method forms the basis of several of the solution procedures that will be elaborated in subsequent chapters. Our starting point is the following assumption.

(A3) The distribution μ of ξ is discrete and has finite support, say $\Xi = \{\xi^1, \dots, \xi^S\}$ with corresponding probabilities $p^1, \dots, p^S$.

For each $s \in \{1, \dots, S\}$, the outcome of random parameters $\big(q(\xi^s), h(\xi^s), T(\xi^s)\big)$, corresponding to the elementary event $\xi^s \in \Xi$, is referred to as a *scenario*, and we denote it simply by (q^s, h^s, T^s).

Remark 2.3.1. Note that (A3) may be justified by the stability results presented in the previous section. In particular, suppose that we are considering a stochastic program with an absolutely continuous distribution of random parameters, rendering numerical computations intractable. Under the assumptions of Theorem 2.2.7, the optimal value and the solution set of this problem may be approximated to any given accuracy by the optimal value and the solution set of problems in which only discrete distributions are employed.

As already pointed out, the fundamental idea underlying the L-shaped algorithm is to perform an outer linerization of the problem. To formalize matters, we write the stochastic program (2.1.1) in the following equivalent form,

$$\begin{aligned}
\min\ & cx + \theta, && (2.3.1a)\\
\text{s.t. } & Ax = b, && (2.3.1b)\\
& \theta \geq \mathcal{Q}(x, \mu), && (2.3.1c)\\
& x \in K_2, && (2.3.1d)\\
& x \in \mathbb{R}^{n_1}_+,\ \theta \in \mathbb{R}. && (2.3.1e)
\end{aligned}$$

The algorithm starts from a relaxation of this problem, referred to as the master problem, in which the constraints (2.3.1c) and (2.3.1d) have been removed. Now, the algorithm progresses by alternatingly solving the master problem and adding so-called *feasibility cuts*, representing the polyhedral domain K_2 (cf. Theorem 2.2.1 and Theorem 2.2.2), and *optimality cuts*, representing the convex piecewise linear expected recourse function $\mathcal{Q}$ (cf. Theorem 2.2.3). This approach is formalized in the following.

Recall that $x \in \mathbb{R}_+^{n_1}$ is said to satisfy all induced constraints if $\mathcal{Q}(x, \mu) < \infty$, and under assumption (A3) this is the case if $x \in K_2$ where

$$K_2 = \{x \in \mathbb{R}_+^{n_1} \mid h^s - T^s x \in \operatorname{pos} W,\ s = 1, \ldots, S\}.$$

The feasibility cuts, representing the constraint $x \in K_2$, are derived from a number of linear programming problems, defined for $s \in \{1, \ldots, S\}$ and for some $x \in \mathbb{R}_+^{n_1}$ by

$$\begin{aligned} &\min ev^+ + ev^-, \\ &\text{ s.t. } Wy + Iv^+ - Iv^- = h^s - T^s x, \\ &\qquad y \in \mathbb{R}_+^{n_2},\ v^+, v^- \in \mathbb{R}_+^{m_2}, \end{aligned} \tag{2.3.2}$$

where $e = (1, \ldots, 1) \in \mathbb{R}^{m_2}$, and I is the $m_2 \times m_2$-identity matrix. Problem (2.3.2) is obviously feasible and bounded below by zero for all $x \in \mathbb{R}_+^{n_1}$, and hence an optimal solution always exists. Moreover, the optimal value of the problem is strictly greater than zero if and only if $h^s - T^s x \notin \operatorname{pos} W$. Now, suppose that in some iteration ν of the algorithm, the solution x^ν of the master problem does not satisfy all induced constraints. Then we see that a dual solution $\sigma^\nu \in \{\sigma \in \mathbb{R}^{m_2} \mid \sigma W \leq 0,\ -e \leq \sigma \leq e\}$ exists, such that $\sigma^\nu(h^s - T^s x^\nu) > 0$ for some scenario $s \in \{1, \ldots, S\}$. Furthermore, since for any $x \in K_2$ the optimal value of problem (2.3.2) is equal to zero, we have for all $x \in K_2$ that

$$\sigma^\nu(h^s - T^s x) \leq 0 \tag{2.3.3}$$

since σ^ν is feasible for the dual problem. The linear inequality (2.3.3) is referred to as a feasibility cut. The inequality may be added to the master problem to cut off the current solution $x^\nu \notin K_2$.

Now, suppose that in some iteration ν of the algorithm we have $x^\nu \in K_2$. Then we may proceed to evaluate the expected recourse function $\mathcal{Q}(x^\nu, \mu)$. To this end we solve the second-stage problems, defined for $s \in \{1, \ldots, S\}$ and for some $x \in \mathbb{R}_+^{n_1}$ by

$$\begin{aligned} &\min q^s y, \\ &\text{ s.t. } Wy = h^s - T^s x, \\ &\qquad y \in \mathbb{R}_+^{n_2}. \end{aligned} \tag{2.3.4}$$

For $s \in \{1, \ldots, S\}$, we let $\pi^{s,\nu}$ be an optimal dual solution of problem (2.3.4) with $x = x^\nu$. By linear programming duality, we have for all $x \in \mathbb{R}^{n_1}$ and $s \in \{1, \ldots, S\}$ that

$$\Phi(x, \xi^s) \geq \pi^{s,\nu}(h^s - T^s x) \tag{2.3.5}$$

with equality holding for $x = x^\nu$, and hence we have for all $x \in \mathbb{R}^{n_1}$ that

$$\mathcal{Q}(x, \mu) \geq \sum_{s=1}^{S} p^s \pi^{s,\nu}(h^s - T^s x)$$

with equality holding for $x = x^\nu$. Therefore, if the solution of the master problem (x^ν, θ^ν) is such that $\theta^\nu < \mathcal{Q}(x^\nu, \mu)$, we may cut off the current solution by adding the following optimality cut to the master problem,

$$\theta \geq \sum_{s=1}^{S} p^s \pi^{s,\nu}(h^s - T^s x). \tag{2.3.6}$$

The original L-shaped algorithm, introduced by Van Slyke and Wets [159], may now be stated in detail as follows.

Algorithm 2.1 *(L-shaped Decomposition)*

Step 1 (*Initialization*) Let $K > 0$, set $\nu = 0$ and $\bar{z} = \infty$, and let the current master problem be $\min\{cx + \theta \mid Ax = b,\ x \in \mathbb{R}^{n_1}_+,\ \theta \in \mathbb{R}\}$.

Step 2 (*Solve master problem*) Set $\nu = \nu + 1$. Solve the current master problem and let (x^ν, θ^ν) be an optimal solution vector if one exists; if the problem is unbounded, then let (x^ν, θ^ν) be a feasible solution with $cx^\nu + \theta^\nu < \bar{z} - K$

Step 3 (*Termination*) If $cx^\nu + \theta^\nu = \bar{z}$, stop; the current solution is optimal.

Step 4 (*Add feasibility cuts*) For each $s \in \{1, \ldots, S\}$, solve the phase-one problem (2.3.2) with $x = x^\nu$ and let $\sigma^{s,\nu}$ be a corresponding optimal dual solution. If $\sigma^{s,\nu}(h^s - T^s x^\nu) > 0$ for some $s \in \{1, \ldots, S\}$, add a feasibility cut (2.3.3) to the master problem and return to Step 2.

Step 5 (*Add optimality cuts*) For each $s \in \{1, \ldots, S\}$, solve the second-stage problem (2.3.4) with $x = x^\nu$ and let $\pi^{s,\nu}$ be a corresponding optimal dual solution. If $\theta^\nu < \sum_{s=1}^{S} p^s \pi^{s,\nu}(h^s - T^s x^\nu)$, add an optimality cut (2.3.6) to the master problem and return to Step 2.

Step 6 (*Update bound*) Let $\bar{z} = \min\{\bar{z},\ cx^\nu + \theta^\nu\}$. Go to Step 2.

Finite convergence of the L-shaped algorithm is a consequence of the fact that only a finite number of feasibility cuts and optimality cuts can be generated since only a finite number of different optimal dual solutions of problems (2.3.2) and (2.3.4) exist. Thus, we have the following.

Proposition 2.3.1. *Assume (A3). If the problem (2.1.1) is feasible and bounded, then Algorithm 2.1 terminates with an optimal solution in a finite number of iterations.*

2.3.2 Extensions

Consider now the following alternative reformulation of problem (2.1.1),

$$
\begin{aligned}
\min\ & cx + \sum_{s=1}^{S} \theta^s \\
\text{s.t. }\ & Ax = b, \\
& \theta^s \geq p^s \Phi(x, \xi^s), \quad s = 1, \ldots, S, \\
& x \in K_2 \\
& x \in \mathbb{R}^{n_1}_+,\ \theta^1, \ldots, \theta^S \in \mathbb{R}.
\end{aligned}
$$

This reformulation leads directly to a multicut version of Algorithm 2.1, presented by Birge and Louveaux [21]. Here the aggregate optimality cut (2.3.6) is replaced by separate cuts on the second-stage value functions $\Phi(\cdot, \xi^s)$ for $s \in \{1, \ldots, S\}$. Thus, let $(x^\nu, \theta^{1,\nu}, \ldots, \theta^{S,\nu})$ be a solution in some iteration ν such that $x^\nu \in K_2$, and for some scenario $s \in \{1, \dots, S\}$ let $\pi^{s,\nu}$ be an optimal dual solution of the second-stage problem (2.3.4) with $x = x^\nu$ such that $\theta^{s,\nu} < p^s \pi^{s,\nu}(h^s - T^s x^\nu)$. Then we may cut off the current solution by adding to the master problem the disaggregate optimality cut,

$$\theta^s \geq p^s \pi^{s,\nu}(h^s - T^s x), \tag{2.3.7}$$

which is valid for all $x \in \mathbb{R}^{n_1}$ cf. (2.3.5). Hence the multicut version of the L-shaped algorithm proceeds as follows.

Algorithm 2.2 *(Multicut L-shaped Decomposition)*

Step 1 (*Initialization*) Let $K > 0$, set $\nu = 0$ and $\bar{z} = \infty$, and let the current master problem be $\min\bigl\{cx + \sum_{s=1}^{S} \theta^s \mid Ax = b,\ x \in \mathbb{R}_+^{n_1},\ \theta^1, \dots, \theta^S \in \mathbb{R}\bigr\}$.

Step 2 (*Solve master problem*) Set $\nu = \nu + 1$. Solve the current master problem and let $(x^\nu, \theta^{1,\nu}, \dots, \theta^{S,\nu})$ be an optimal solution vector if one exists; if the problem is unbounded, then let $(x^\nu, \theta^{1,\nu}, \dots, \theta^{S,\nu})$ be a feasible solution with $cx^\nu + \sum_{s=1}^{S} \theta^{s,\nu} < \bar{z} - K$.

Step 3 (*Termination*) If $cx^\nu + \sum_{s=1}^{S} \theta^{s,\nu} = \bar{z}$, stop; the current solution is optimal.

Step 4 (*Add feasibility cuts*) For each $s \in \{1, \dots, S\}$, solve the phase-one problem (2.3.2) with $x = x^\nu$ and let $\sigma^{s,\nu}$ be a corresponding optimal dual solution. If $\sigma^{s,\nu}(h^s - T^s x^\nu) > 0$ for some $s \in \{1, \dots, S\}$, add a feasibility cut (2.3.3) to the master problem and return to Step 2.

Step 5 (*Add optimality cuts*) For each $s \in \{1, \dots, S\}$, solve the second-stage problem (2.3.4) with $x = x^\nu$ and let $\pi^{s,\nu}$ be a corresponding optimal dual solution. If $\theta^{s,\nu} < p^s \pi^{s,\nu}(h^s - T^s x^\nu)$ for some $s \in \{1, \dots, S\}$, add an optimality cut (2.3.7) to the master problem and return to Step 2.

Step 6 (*Update bound*) Let $\bar{z} = \min\bigl\{\bar{z},\ cx^\nu + \sum_{s=1}^{S} \theta^{s,\nu}\bigr\}$. Go to Step 2.

Obviously, we have the following equivalent to Proposition 2.3.1.

Proposition 2.3.2. *Assume (A3). If the problem (2.1.1) is feasible and bounded, then Algorithm 2.2 terminates with an optimal solution in a finite number of iterations.*

The multicut approach of Algorithm 2.2 offers some potential computational advantages compared to Algorithm 2.1, since more detailed information is passed to the master problem in each iteration. Therefore, the number of overall iterations required by Algorithm 2.2 is in general expected to be less than required by Algorithm 2.1. The improved detailing, however, comes at the cost of an increased complexity of the master problem since the size of the problem increases more rapidly. This is particularly a problem because no reliable criterion for removing inactive cuts from the master problem exists for any of these procedures. Apart from the growing size of the master problem, a major drawback of the L-shaped algorithm and its multicut extension is the tendency for early iterations to oscillate heavily, causing slow convergence toward an optimal solution. These drawbacks were circumvented in the regularized decomposition method introduced by Ruszczyński [132]. The algorithm is based on ideas known from *bundle methods* for non-smooth optimization. (See e.g. Kiwiel [71].) Here, an incumbent solution a^ν is introduced, and a quadratic regularizing term of the form $\frac{1}{2}\|x - a^\nu\|^2$ is included in the objective of the master problem. Hence the master problem in iteration ν is stated as,

$$\begin{aligned}
&\min cx + \sum_{s=1}^{S} \theta^s + \frac{1}{2}\|x - a^\nu\|^2 \\
&\text{s.t. } Ax = b, \\
&\qquad \sigma^{s,l}(h^s - T^s x) \le 0, \qquad l \in \mathcal{I}^{s,\nu},\ s = 1, \ldots, S, \\
&\qquad \theta^s \ge p^s \pi^{s,l}(h^s - T^s x), \quad l \in \mathcal{J}^{s,\nu},\ s = 1, \ldots, S, \\
&\qquad x \in \mathbb{R}^{n_1}_+,\ \theta^1, \ldots, \theta^S \in \mathbb{R}.
\end{aligned} \tag{2.3.8}$$

where $\mathcal{I}^{s,\nu}$ and $\mathcal{J}^{s,\nu}$ are index sets for the feasibility cuts and optimality cuts, respectively, that are present in the master problem in iteration ν for scenario $s \in \{1, \ldots, S\}$. Now, the regularized decomposition algorithm may be stated as follows.

Algorithm 2.3 *(Regularized Decomposition)*

Step 1 (*Initialization*) Set $\nu = 0$, $\bar{z} = \infty$, and $\mathcal{I}^{s,\nu} = \mathcal{J}^{s,\nu} = \emptyset$ for $s = 1, \ldots, S$. Choose a starting point a^1 and let the master problem be defined by (2.3.8).

Step 2 (*Solve master problem*) Set $\nu = \nu + 1$. Solve the current master problem and let $(x^\nu, \theta^{1,\nu}, \ldots, \theta^{S,\nu})$ be an optimal solution vector. (If $\mathcal{J}^{s,\nu} = \emptyset$ for some $s \in \{1, \ldots, S\}$, the corresponding variable θ^s is ignored in the computation.)

Step 3 (*Termination*) If $cx^\nu + \sum_{s=1}^{S} \theta^{s,\nu} = \bar{z}$, stop; the incumbent a^ν is optimal.

Step 4 (*Add feasibility cuts*) For each $s \in \{1, \ldots, S\}$, solve the phase-one problem (2.3.2) with $x = x^\nu$ and let $\sigma^{s,\nu}$ be a corresponding optimal dual solution. If $\sigma^{s,\nu}(h^s - T^s x^\nu) > 0$ for some $s \in \{1, \ldots, S\}$, then let $\mathcal{I}^{s,\nu+1} = \mathcal{I}^{s,\nu} \cup \{\nu\}$, set $a^{\nu+1} = a^\nu$, and go to Step 9 (*null infeasible step*).

Step 5 (*Add optimality cuts*) For each $s \in \{1, \ldots, S\}$, solve the second-stage problem (2.3.4) with $x = x^\nu$ and let $\pi^{s,\nu}$ be a corresponding optimal dual solution. If $\theta^{s,\nu} < p^s \pi^{s,\nu}(h^s - T^s x^\nu)$ for some $s \in \{1, \ldots, S\}$, then let $\mathcal{J}^{s,\nu+1} = \mathcal{J}^{s,\nu} \cup \{\nu\}$ and go to Step 7.

Step 6 (*Exact serious step*) Set $\bar{z} = cx^\nu + \sum_{s=1}^{S} \theta^{s,\nu}$ and $a^{\nu+1} = x^\nu$. Go to step 9.

Step 7 (*Approximate serious step*) If $cx^\nu + \sum_{s=1}^{S} p^s \pi^{s,\nu}(h^s - T^s x^\nu) < \bar{z}$, then set $\bar{z} = cx^\nu + \sum_{s=1}^{S} p^s \pi^{s,\nu}(h^s - T^s x^\nu)$ and $a^{\nu+1} = x^\nu$, and go to step 9.

Step 8 (*Null feasible step*) Set $a^{\nu+1} = a^\nu$.

Step 9 (*Cut deletion*) For $s \in \{1, \ldots, S\}$, remove from $\mathcal{I}^{s,\nu}$ and $\mathcal{J}^{s,\nu}$ some indices, corresponding to cuts that were inactive at the solution $(x^\nu, \theta^{1,\nu}, \ldots, \theta^{S,\nu})$, to obtain $\mathcal{I}^{s,\nu+1}$ and $\mathcal{J}^{s,\nu+1}$, respectively. Go to step 2.

Once again we have the following.

Proposition 2.3.3. *Assume (A3). If the problem (2.1.1) is feasible and bounded, then Algorithm 2.3 terminates with an optimal solution in a finite number of iterations.*

Chapter 3

Two-Stage Stochastic Programs with Mixed-Integer Recourse

Many real-life decision problems involve decisions that are by nature discrete. In fact, all of the applications of stochastic programming considered in Chapters 6 through 10 of this thesis, lead to models with either binary or general integer variables. When integer variables occur only in the first stage of a two-stage stochastic program with recourse, the difficulties are not too severe, since the structural properties of the expected recourse function established in Chapter 2 remain unchanged. Examples of such models will be considered in Chapters 7 and 10. When integer variables occur in the second stage, on the other hand, matters are seriously complicated since the expected recourse function is no longer necessarily convex nor even continuous. Examples of recourse models with (mixed-) integer second stage will be considered in Chapters 6 and 9. In this chapter we consider a general two-stage stochastic program with mixed-integer recourse. We survey results on structural properties related to continuity of the expected recourse function and to stability of optimal solutions when the underlying probability measure is subjected to perturbations. Also, we discuss a number of alternative decomposition-based solution procedures for the problem. For a recent survey of research in the field of stochastic integer programming, we refer to Klein Haneveld and van der Vlerk [79]. See also the annotated bibliography by Stougie and van der Vlerk [151], and the extensive online bibliography of van der Vlerk [157].

3.1 Problem Formulation

We recall from Section 1.2.2 that the classical two-stage stochastic program with recourse concerns minimization of the sum of direct cost and expected recourse cost arising from some first-stage decision. This first-stage decision must be made without certain knowledge on some random parameters of the model, whereas the recourse cost is determined as a second-stage decision is made after uncertainty has been disclosed. As mentioned above, we assume throughout this chapter that some of the second-stage variables are restricted to integer values. Structural properties of the problem in this case have been established only under the assumption that the recourse matrix and the second-stage cost are fixed, and hence we employ this assumption throughout the structural analysis.

As in Chapter 2 the stability analysis of optimal solutions will require continuity properties of the expected recourse function as a function of the underlying probability distribution as well as of the first-stage decision, and hence once again we find the following formulation of the problem convenient. We let $\xi : \Omega \mapsto \mathbb{R}^N$ be a random vector defined on some probability space $(\Omega, \mathcal{F}, P)$, the components of which constitute the random second-stage data, consisting of the second-stage right-hand side $\tilde{h} : \Omega \mapsto \mathbb{R}^{m_2}$ and the technology matrix $\tilde{T} : \Omega \mapsto \mathbb{R}^{m_2 \times n_1}$. In other words we have $N = m_2(1 + n_1)$, and for $\omega \in \Omega$ we have $\xi(\omega) = \big(\tilde{h}(\omega), \tilde{T}_1(\omega), \ldots, \tilde{T}_{m_2}(\omega)\big)$ where $\tilde{T}_k$ denotes the kth row of $\tilde{T}$ for $k = 1, \ldots, m_2$. Introducing the induced Borel probability measure $\mu = P \circ \xi^{-1}$ on $\mathbb{R}^N$, the two-stage mixed-integer recourse problem is now formulated as,

$$TSMIR(\mu) \qquad \min\big\{cx + \mathcal{Q}(x, \mu) \mid Ax = b,\ x \in X\big\}, \tag{3.1.1}$$

where the expected recourse function $\mathcal{Q}$ is given by,

$$\mathcal{Q}(x, \mu) = \mathbb{E}\big[\Phi\big(h(\xi) - T(\xi)x\big)\big] = \int_{\mathbb{R}^N} \Phi\big(h(\xi) - T(\xi)x\big)\mu(d\xi), \tag{3.1.2}$$

and the second-stage value function Φ is given by

$$\Phi(\tau) = \min\big\{qy + q'y' \mid Wy + W'y' = \tau,\ y \in \mathbb{Z}_+^{n_2},\ y' \in \mathbb{R}_+^{n_2'}\big\}. \tag{3.1.3}$$

Here $c \in \mathbb{R}^{n_1}$, $q \in \mathbb{R}^{n_2}$, $q' \in \mathbb{R}^{n_2'}$, and $b \in \mathbb{R}^{m_1}$ are known vectors, and $A \in \mathbb{R}^{m_1 \times n_1}$, $W \in \mathbb{R}^{m_2 \times n_2}$, and $W' \in \mathbb{R}^{m_2 \times n_2'}$ are known matrices. Moreover, it is assumed that W and W' have rational entries. The second-stage right-hand side $h : \mathbb{R}^N \mapsto \mathbb{R}^{m_2}$ and the technology matrix $T : \mathbb{R}^N \mapsto \mathbb{R}^{m_2 \times n_1}$, on the other hand, are represented as mappings picking out the appropriate components of the random vector ξ. The set $X \subseteq \mathbb{R}_+^{n_1}$ is assumed to be non-empty and closed, and in particular it may or may not contain integrality restrictions on some or all of the first-stage variables.

In the following we let $\mu \in \mathcal{P}(\mathbb{R}^N)$ represent the joint distribution of the second-stage right-hand side $\tilde{h}$ and the technology matrix $\tilde{T}$ as described above. At some points we will also be interested in the marginal distributions of $\tilde{h}$ and $\tilde{T}$, and we denote these by μ_1 and μ_2, respectively. Moreover, for $T \in \mathbb{R}^{m_2 \times n_1}$ we denote by $\mu_1^2(\cdot, T)$ the (regular) conditional distribution of $\tilde{h}$ given $\tilde{T} = T$. (See Appendix A.)

Remark 3.1.1. In Remark 2.1.2 on page 15 we briefly mentioned the important special case of two-stage stochastic programs with linear recourse referred to as simple recourse problems. Clearly, this notion may be generalized for problems with integer recourse. In particular, consider the special case of the stochastic program (3.1.1)-(3.1.2) where the second-stage value function is given by,

$$\Phi(\tau) = \min\big\{q^+y^+ + q^-y^- \mid y^+ \geq \tau,\ y^- \geq -\tau,\ y^+, y^- \in \mathbb{Z}_+^{m_2}\big\},$$

where $q^+, q^- \in \mathbb{R}^{m_2}$ are known vectors. This is referred to as the simple integer recourse case. As in the linear case, stochastic programs with simple integer recourse have received particular attention in the literature, since the special structure of the problem provides significant computational advantages. Hence, structural properties preparing the way for efficient solution procedures for the simple integer recourse problem, have been established in papers by Klein Haneveld, Stougie, and van der Vlerk [74, 75, 76, 77], Klein Haneveld and van der Vlerk [78], and Louveaux and van der Vlerk [86]. See also the PhD-thesis by van der Vlerk [156].

3.2 Structural Properties

As already pointed out, the complications associated with two-stage stochastic programs with mixed-integer recourse arise mainly due to the fact that properties of the second-stage value function such as continuity and convexity are lost when integer requirements are imposed on second-stage variables. Therefore, to establish results related to continuity properties of the expected recourse function and to stability properties of optimal solutions, the assumptions employed during the structural analysis of linear recourse problems must be complemented with further restrictions on the probability measure to ensure a sufficient "smoothing effect" of the integral in (3.1.2). An early result in this respect, concerning continuity of the expected recourse function, was established by Stougie [150] (see also Rinnooy Kan and Stougie [119]). Furthermore, Artstein and Wets [4] considered maximization of an integral functional with discontinuous integrands in a general setting, containing the two-stage stochastic program with mixed-integer recourse as a special case. In this setting the authors study lower and upper semicontinuity of the integral functional, and establish qualitative stability results of the problem, similar in vein to those presented for the mixed-integer recourse problem in this section. Apart from these references, most results on structural properties of two-stage stochastic programs with mixed-integer recourse are due to Schultz [138, 139, 140, 141].

3.2.1 The Expected Recourse Function

During the survey of structural properties for two-stage stochastic programs with linear recourse, presented in Section 2.2, we employed the assumptions of complete recourse and dual feasibility of the second-stage problems cf. (A1) and (A2) on page 17. These assumptions were complemented with different moment conditions to establish the desired structural properties of the expected recourse function. In this section we will employ the following counterparts of (A1) and (A2).

(B1) For all $t \in \mathbb{R}^{m_2}$ there exist $y \in \mathbb{Z}_+^{n_2}$ and $y' \in \mathbb{R}_+^{n_2'}$ such that $Wy + W'y' = t$.

(B2) There exists $u \in \mathbb{R}^{m_2}$ such that $uW \leq q$ and $uW' \leq q'$.

Here (B1) is the natural extension of the complete recourse property for stochastic programs with linear recourse (cf. Definition 2.2.2), and hence it is referred to as the complete mixed-integer recourse property. Assumption (B2), on the other hand, implies dual feasibility of the linear relaxation of the second-stage problem. These two assumptions ensure feasibility and boundedness, respectively, of the second-stage problem, and hence they are sufficient to establish the following properties of the second-stage value-function defined by (3.1.3). For these and related results on the value function of a mixed-integer program, we refer to e.g. the monograph by Bank and Mandel [10] and the research paper by Blair and Jeroslow [24].

Lemma 3.2.1. *Assume (B1)-(B2). Then Φ is a real-valued lower semicontinuous function on $\mathbb{R}^{m_2}$ and*

(a) there exist constants $\alpha > 0$ and $\beta > 0$ such that for all $\tau, \tau' \in \mathbb{R}^{m_2}$ we have $\left|\Phi(\tau) - \Phi(\tau')\right| \leq \alpha||\tau - \tau'|| + \beta$;

(b) there exist constants $\gamma > 0$ and $\delta > 0$ and vectors $d_1, \ldots, d_l, \tilde{d}_1, \ldots, \tilde{d}_{l'} \in \mathbb{R}^{m_2}$ such that for all $\tau \in \mathbb{R}^{m_2}$ we have $\Phi(\tau) = \min\{qy + \max_{k\in\{1,\ldots,l\}} d_k(\tau - Wy) \mid y \in Y(\tau)\}$ where $Y(\tau) = \{y \in \mathbb{Z}_+^{n_2} \mid \sum_i |y_i| \leq \gamma \sum_j |\tau_j| + \delta,\ \tilde{d}_k(\tau - Wy) \geq 0,\ k = 1, \ldots, l'\}$.

To arrive at the desired properties of the expected recourse function, (B1) and (B2) must be complemented with the additional assumption that the probability distribution has finite first moment, i.e. we employ the moment condition $\mu \in \mathcal{M}_1(\mathbb{R}^N)$, where we recall that

$$\mathcal{M}_1(\mathbb{R}^N) = \left\{\mu \in \mathcal{P}(\mathbb{R}^N) \;\Big|\; \int_{\mathbb{R}^N} ||\xi||\mu(d\xi) < \infty\right\}.$$

With the result of Lemma 3.2.1 (a) this restriction on the probability measure is sufficient to establish the existence of an integrable minorant of the second-stage value function, and hence Schultz [140] applied Fatou's Lemma to arrive at the following result.

Theorem 3.2.1. *Assume (B1)-(B2) and let $\mu \in \mathcal{M}_1(\mathbb{R}^N)$. Then $\mathcal{Q}(\cdot, \mu)$ is a real-valued lower semicontinuous function on $\mathbb{R}^{n_1}$.*

Note that with fixed second-stage cost, the condition $\mu \in \mathcal{M}_1(\mathbb{R}^N)$ is sufficient to ensure that the weak covariance condition is satisfied (cf. Definition 2.2.1), and hence the assumptions of Theorem 3.2.1 closely correspond to those of Theorem 2.2.3 on page 18, where (Lipschitz) continuity of the expected recourse function was stated for problems with linear recourse. To arrive at sufficient conditions for continuity of the expected recourse function for problems with mixed-integer recourse, however, further assumptions must be made about the probability measure μ. To this end we define for some $x \in \mathbb{R}^{n_1}$ the set $E(x)$ of all those $\xi \in \mathbb{R}^N$ such that $h(\xi) - T(\xi)x$ is a discontinuity point of Φ. Using again the result of Lemma 3.2.1 (a) to establish the existence of an integrable majorant of the second-stage value function, and applying Lebesgue's dominated convergence theorem, Schultz [140] now showed the following.

Theorem 3.2.2. *Assume (B1)-(B2) and let $\mu \in \mathcal{M}_1(\mathbb{R}^N)$ and $x \in \mathbb{R}^{n_1}$ be such that $\mu(E(x)) = 0$. Then $\mathcal{Q}(\cdot, \mu)$ is continuous at x.*

Furthermore, applying the result of Lemma 3.2.1 (b), it may be seen that the set of all discontinuity points of Φ is contained in a countable union of hyperplanes in $\mathbb{R}^{m_2}$. This leads directly to the following corollary.

Corollary 3.2.1. *Assume (B1)-(B2) and let $\mu \in \mathcal{M}_1(\mathbb{R}^N)$ be such that $\mu_1^2(\cdot, T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$ for μ_2-almost all $T \in \mathbb{R}^{m_2 \times n_1}$. Then $\mathcal{Q}(\cdot, \mu)$ is a continuous function on $\mathbb{R}^{n_1}$.*

Finally, Schultz [140] also established a sufficient condition for Lipschitz continuity of the expected recourse function.

Theorem 3.2.3. *Adopt the setting of Corollary 3.2.1, and assume further that for any non-singular transformation $B : \mathbb{R}^{m_2} \mapsto \mathbb{R}^{m_2}$ and for μ_2-almost all $T \in \mathbb{R}^{m_2 \times n_1}$, the one-dimensional marginal distributions of $\mu_1^2(\cdot, T) \circ B$ have densities that are uniformly bounded with respect to T and monotonic outside some bounded interval not depending on T. Then $\mathcal{Q}(\cdot, \mu)$ is Lipschitz continuous on any bounded subset of $\mathbb{R}^{n_1}$.*

Remark 3.2.1. The assumptions of Theorem 3.2.3 may seem rather technical. However, they may be seen to hold for example for the (non-degenerate) multivariate normal distribution and for t-distributions. Furthermore, Schultz [139] provided two examples of integer recourse problems with one-dimensional random right-hand side, showing that the Lipschitz continuity of the expected recourse function may fail if either of the assumptions on boundedness or monotonicity in Theorem 3.2.3 are abandoned.

Next we consider results related to joint continuity of $\mathcal{Q}$ as a function of the first-stage decision and the underlying probability measure. As in the linear recourse case, we recall that the set of Borel probability measures is endowed with the notion of weak convergence (see Appendix A), and note that the continuity of $\mathcal{Q}$ with respect to μ requires some uniform integrability condition to be satisfied cf. e.g. Billingsley [18, Theorem 5.4] or Hoffman-Jørgensen [61, Section 5.2]. For mixed-integer recourse problems, Schultz [140] considered the family of Borel probability measures given by

$$\mathcal{P}_{p,K}(\mathbb{R}^N) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^N) \;\Big|\; \int_{\mathbb{R}^N} ||\xi||^p \mu(d\xi) \leq K \right\},$$

for some real numbers $p > 1$ and $K > 0$, and showed that Theorem 5.4 and Theorem 5.5 in Billingsley [18] can be applied to obtain the following.

Theorem 3.2.4. *Assume (B1)-(B2) and for some $p > 1$ and $K > 0$ let $\mu \in \mathcal{P}_{p,K}(\mathbb{R}^N)$ and $x \in \mathbb{R}^{n_1}$ be such that $\mu(E(x)) = 0$. Then $\mathcal{Q}$, as a function from $\mathbb{R}^{n_1} \times \mathcal{P}_{p,K}(\mathbb{R}^N)$ to $\mathbb{R}$, is continuous at (x, μ).*

Again we have the following immediate corollary.

Corollary 3.2.2. *Assume (B1)-(B2) and for some $p > 1$ and $K > 0$ let $\mu \in \mathcal{P}_{p,K}(\mathbb{R}^N)$ be such that $\mu_1^2(\cdot, T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$ for μ_2-almost all $T \in \mathbb{R}^{m_2 \times n_1}$. Then $\mathcal{Q}$, as a function from $\mathbb{R}^{n_1} \times \mathcal{P}_{p,K}(\mathbb{R}^N)$ to $\mathbb{R}$, is continuous on $\mathbb{R}^{n_1} \times \{\mu\}$.*

Remark 3.2.2. Once again note that with fixed second-stage cost, the assumptions of Theorem 3.2.4 and Corollary 3.2.2 correspond to those of Theorem 2.2.4 on page 18, complemented with an additional restriction on the probability measure, required to ensure a sufficient "smoothing effect" of the integral for the mixed-integer recourse case.

As for quantitative continuity results of the expected recourse function with respect to the underlying probability measure, Schultz [140] gave an example, showing that no Hölder estimate for $\mathcal{Q}(x, \cdot)$ with respect to the L_1-Wasserstein metric nor the bounded Lipschitz metric can be established when integer requirements are imposed on the second-stage variables. Hence the challenge has been to come up with a suitable distance of probability measures that fits the discontinuities of the second-stage value function and at the same time covers weak convergence of probability measures. First results in this direction were obtained by Schultz [141] for problems with a fixed technology matrix. Hence we assume in the following that only the second-stage right-hand side is random and thus consider probability measures in $\mathcal{P}(\mathbb{R}^{m_2})$. Schultz [141] considered the following variational distance (or discrepancy), defined for $\mu, \nu \in \mathcal{P}(\mathbb{R}^{m_2})$ by

$$\alpha_{\mathcal{B}_\mathcal{K}}(\mu, \nu) = \sup\big\{|\mu(B) - \nu(B)| \;\big|\; B \in \mathcal{B}_\mathcal{K}\big\}, \tag{3.2.1}$$

where $\mathcal{B}_{\mathcal{K}} \subseteq \mathcal{B}(\mathbb{R}^{m_2})$ denotes the class of all (closed) bounded polyhedra in $\mathbb{R}^{m_2}$ whose facets parallels a facet of $\mathcal{K} = \operatorname{pos} W'$ or a facet of the m_2-dimensional unit hyper-cube (i.e. $\times_{i=1}^{m_2}[0,1]$). Using this discrepancy, Schultz was able to obtain the following Hölder estimate for the expected recourse function with respect to the distribution of the second-stage right-hand side.

Theorem 3.2.5. *Let the technology matrix be fixed and assume (B1)-(B2). Furthermore, let $D \subseteq \mathbb{R}^{n_1}$ be non-empty and compact, and let $\mu \in \mathcal{P}_{p,K}(\mathbb{R}^{m_2})$ for some $p > 1$ and $K > 0$. Then there exist $L > 0$ and $\delta > 0$ such that*

$$\big|\mathcal{Q}(x,\mu) - \mathcal{Q}(x,\nu)\big| \leq L \cdot \alpha_{\mathcal{B}_{\mathcal{K}}}(\mu,\nu)^{\frac{p-1}{p(m_2+1)}}$$

for all $x \in D$ and all $\nu \in \mathcal{P}_{p,K}(\mathbb{R}^{m_2})$ with $\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu,\nu) < \delta$.

Remark 3.2.3. Adapting probability (pseudo-) distances to the underlying structures is a proven tool in quantitative stability analysis of stochastic programs. We refer to Rachev and Römisch [109] for a general framework and applications to classical recourse models as well as chance-constrained problems.

Remark 3.2.4. Schultz [141] showed that the discrepancy $\alpha_{\mathcal{B}_{\mathcal{K}}}$ defined above is in fact a metric on $\mathcal{P}(\mathbb{R}^{m_2})$. Moreover, he used the concept of a μ-uniformity class to establish a coherence between the discrepancy $\alpha_{\mathcal{B}_{\mathcal{K}}}$ and weak convergence of probability measures as follows. For $\mu \in \mathcal{P}(\mathbb{R}^{m_2})$, a family of Borel sets $\mathcal{B}_0 \subseteq \mathcal{B}(\mathbb{R}^{m_2})$ is called a μ-uniformity class if for every sequence of probability measures $\{\mu_n\}_{n=1}^{\infty}$ in $\mathcal{P}(\mathbb{R}^{m_2})$ converging weakly to μ we have $\sup\big\{|\mu_n(B) - \mu(B)| \;\big|\; B \in \mathcal{B}_0\big\} \xrightarrow{n\to\infty} 0$. Now, according to Bhattacharya and Ranga Rao [15, Theorem 2.11], the class $\mathcal{B}_c$ of all convex Borel sets in $\mathbb{R}^{m_2}$ is a μ-uniformity class for all those $\mu \in \mathcal{P}(\mathbb{R}^{m_2})$ that are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$. Therefore, since we obviously have $\mathcal{B}_{\mathcal{K}} \subseteq \mathcal{B}_c$, we see that $\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu_n,\mu) \xrightarrow{n\to\infty} 0$ for any sequence of probability measures $\{\mu_n\}_{n=1}^{\infty}$ in $\mathcal{P}(\mathbb{R}^{m_2})$ converging weakly to such μ.

3.2.2 Stability

Once again let us note that in many practical applications of stochastic programming, the probability distribution of random parameters will not be completely known, and hence the true distribution may have to be replaced in the model by some suitable estimate, such as e.g. empirical measures. Furthermore, even if the true distribution μ of random parameters is known, the approximation of μ by simpler probability measures may be required to facilitate practical computations. In fact, as we will see in Section 3.3, most solution procedures for two-stage stochastic programs with (mixed-) integer recourse rely on the assumption that the probability distribution of random parameters is discrete with finite support. Obviously, though, this will quite often not be the case in practice. Therefore, just as in the linear recourse case, stability of the two-stage stochastic program with mixed-integer recourse $TSMIR(\mu)$, when the underlying probability distribution is subjected to perturbations, is an important issue to which we now turn.

Since the two-stage stochastic program with mixed-integer recourse $TSMIR(\mu)$, defined by (3.1.1)-(3.1.3), is in general a non-convex problem, local minimizers are now included in the analysis. Therefore, the stability results presented here take the form of continuity properties of a localized version of the optimal-value function, $\varphi_V : \mathcal{P}(\mathbb{R}^N) \mapsto \mathbb{R}$, defined for some non-empty open set $V \subseteq \mathbb{R}^{n_1}$ by

$$\varphi_V(\mu) = \inf\{cx + \mathcal{Q}(x, \mu) \mid Ax = b,\ x \in X \cap \operatorname{cl} V\},$$

and of a localized version of the solution set mapping, $\Psi_V : \mathcal{P}(\mathbb{R}^N) \mapsto \mathbb{R}^{n_1}$, that is defined accordingly by

$$\Psi_V(\mu) = \arg\min\{cx + \mathcal{Q}(x, \mu) \mid Ax = b,\ x \in X \cap \operatorname{cl} V\}.$$

where $\operatorname{cl} V$ denotes the closure of V.

Having established the joint continuity of $\mathcal{Q}$ with respect to x and μ, it is straightforward to prove continuity of φ_V and Berge upper semicontinuity of Ψ_V for any bounded open set $V \subseteq \mathbb{R}^{n_1}$. (Once again recall that the point-to-set mapping, Ψ_V, is Berge upper semicontinuous at some $\mu \in \mathcal{P}(\mathbb{R}^N)$ if for any open set $G \subseteq \mathbb{R}^{n_1}$ with $\Psi_V(\mu) \subseteq G$ there exists some neighborhood U of μ in $\mathcal{P}(\mathbb{R}^N)$ such that $\Psi_V(\nu) \subseteq G$ for all $\nu \in U$.) For the analysis of local minimizers, however, we will not find these properties alone quite sufficient, the shortcoming being that they do not preclude certain pathologies that may occur when dealing with stability of local minimizers. The difficulties that may occur are illustrated by the following example.

Example 3.2.1. Consider the two-stage stochastic program with mixed-integer recourse, $\min\{2x + \mathcal{Q}(x, \mu) \mid x \geq 0\}$, where $\mathcal{Q}(x, \mu) = \mathbb{E}_\xi\big[\min\{y \mid y \geq \xi - x,\ y \in \mathbb{Z}\}\big]$, and ξ is a uniformly distributed random variable, $\xi \sim U(-\frac{1}{4}, \frac{1}{4})$. Here the distribution μ of ξ is absolutely continuous with density

$$f_\mu(\xi) = \begin{cases} 2 & \text{for } -\frac{1}{4} \leq \xi \leq \frac{1}{4}\,; \\ 0 & \text{otherwise,} \end{cases}$$

and it is easily seen that we have

$$\mathcal{Q}(x, \mu) = \begin{cases} -n & \text{for } n + \frac{1}{4} \leq x \leq n + \frac{3}{4},\ n \in \mathbb{Z}; \\ -2x + n + \frac{1}{2} & \text{for } n - \frac{1}{4} \leq x \leq n + \frac{1}{4},\ n \in \mathbb{Z}. \end{cases}$$

Now, for $\epsilon > 0$ consider the perturbed distribution μ_ϵ having density

$$f_{\mu_\epsilon}(\xi) = \begin{cases} 2 + \epsilon\xi & \text{for } -\frac{1}{4} \leq \xi \leq \frac{1}{4}\,; \\ 0 & \text{otherwise.} \end{cases}$$

For the perturbed distribution we have

$$\mathcal{Q}(x, \mu_\epsilon) = \begin{cases} -n & \text{for } n + \frac{1}{4} \leq x \leq n + \frac{3}{4},\ n \in \mathbb{Z}; \\ -2x + n + \frac{1}{2} + \epsilon\big(\frac{1}{32} - \frac{1}{2}(x - n)^2\big) & \text{for } n - \frac{1}{4} \leq x \leq n + \frac{1}{4},\ n \in \mathbb{Z}. \end{cases}$$

The objective function of the original problem and of the perturbed problem are both illustrated in Figure 3.1. Now, for each $n \in \mathbb{Z}_+$, the closed interval $[n - 0.1, n + 0.1]$, for example, is a bounded set of local minimizers of the original problem, and hence we should like these sets to behave stably under perturbations of the problem. For any $\epsilon > 0$, however, none of the intervals $[n - 0.1, n + 0.1]$, $n \in \mathbb{Z}_+$, contain any local minimizers of the perturbed problem.

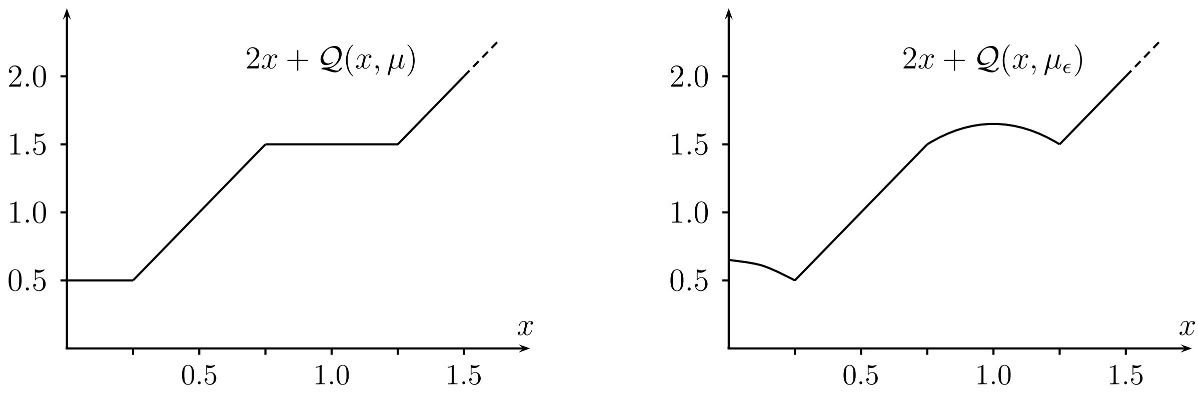

Figure 3.1: Objective function of the original problem and of the perturbed problem in Example 3.2.1.

To preclude pathologies as the one illustrated in Example 3.2.1, Robinson [121] and Klatte [73] proposed a local stability analysis for non-convex problems, emphasizing the need for considerations to include all local minimizers that are, in some sense, nearby the minimizers one is interested in. The crucial concept is that of a complete local minimizing set, or simply a CLM set, which may be formulated as follows. Let μ be a Borel probability measure on $\mathbb{R}^N$, and let M be a non-empty subset of $\mathbb{R}^{n_1}$. If there exists an open set $V \subseteq \mathbb{R}^{n_1}$ such that $M \subseteq V$ and $M = \Psi_V(\mu)$, then M is called a CLM set for $TSMIR(\mu)$ with respect to V. Obvious examples of CLM sets are the set of global minimizers as well as any set of strict local minimizers. Hence, the subsequent results stated in general for CLM sets are valid in particular for the set of global minimizers and for any set of strict local minimizers.

Theorem 3.2.6. *Assume (B1)-(B2), for some $p > 1$ and $K > 0$ let $\mu \in \mathcal{P}_{p,K}(\mathbb{R}^N)$ be such that $\mu_1^2(\cdot, T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$ for μ_2-almost all $T \in \mathbb{R}^{m_2 \times n_1}$, and let $V \subseteq \mathbb{R}^{n_1}$ be a bounded open set such that $\Psi_V(\mu)$ is a CLM set for $TSMIR(\mu)$ with respect to V. Then,*

(a) φ_V as a function from $\mathcal{P}_{p,K}(\mathbb{R}^N)$ to $\mathbb{R}$ is continuous at μ;

(b) Ψ_V as a mapping from $\mathcal{P}_{p,K}(\mathbb{R}^N)$ to $\mathbb{R}^{n_1}$ is Berge upper semicontinuous at μ;

(c) there exists some neighborhood U of μ in $\mathcal{P}_{p,K}(\mathbb{R}^N)$ such that $\Psi_V(\nu)$ is a CLM set for $TSMIR(\nu)$ with respect to V for all $\nu \in U$.

Assuming that the technology matrix is fixed, the result of Theorem 3.2.5 allows the following quantitative stability result.

Theorem 3.2.7. *Let the technology matrix be fixed and assume (B1)-(B2). Furthermore, for some $p > 1$ and $K > 0$ let $\mu \in \mathcal{P}_{p,K}(\mathbb{R}^{m_2})$ be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$, and let $V \subseteq \mathbb{R}^{n_1}$ be a bounded open set such that $\Psi_V(\mu)$ is a CLM set for $TSMIR(\mu)$ with respect to V. Then there exist $L > 0$ and $\delta > 0$ such that*

$$\left|\varphi_V(\mu) - \varphi_V(\nu)\right| \leq L \cdot \alpha_{\mathcal{B}_\mathcal{K}}(\mu, \nu)^{\frac{p-1}{p(m_2+1)}}$$

for all $\nu \in \mathcal{P}_{p,K}(\mathbb{R}^{m_2})$ with $\alpha_{\mathcal{B}_\mathcal{K}}(\mu, \nu) < \delta$.

Remark 3.2.5. Here the assumption that μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$, is required only to ensure that $\mathcal{B}_{\mathcal{K}}$ is a μ-uniformity class cf. Remark 3.2.4, and hence that $\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu_n, \mu) \xrightarrow{n\to\infty} 0$ for any sequence of probability measures $\{\mu_n\}_{n=1}^{\infty}$ in $\mathcal{P}(\mathbb{R}^{m_2})$ converging weakly to μ.

Just as in the linear recourse case we note that the true distribution of random parameters will not be completely known in many practical applications of stochastic programming, and hence the true distribution may have to be replaced by some suitable estimate. Hence we consider again the situation when the true distribution μ is approximated by empirical measures. In particular, we let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of independent and identically distributed N-dimensional random vectors defined on some probability space $(\Omega, \mathcal{F}, P)$, and we denote by μ their common distribution. This gives rise to a corresponding sequence of empirical probability measures on $\mathbb{R}^N$ defined by

$$\mu_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i(\omega)},$$

where $\delta_{\xi_i(\omega)}$ denotes the measure with unit mass at $\xi_i(\omega)$ for $i = 1, \ldots, n$. It is well-known that we have $\mu_n(\omega) \xrightarrow{w} \mu$ for P-almost all $\omega \in \Omega$ cf. e.g. Dudley [42, Theorem 11.4.1]. As in the linear recourse case one may use this fact and apply Theorem 3.2.6 to obtain asymptotic convergence of local optimal values and local optimal solutions when μ satisfies the hypotheses of that theorem, cf. Theorem 2.2.8 on page 21. For problems with a fixed technology matrix, however, Schultz [141], went another way, showing that the smoothness assumption on μ in Theorem 3.2.6 can be abandoned. In particular, Schultz used the concept of a so-called Vapnik-Červonenkis class to prove that for any probability measure $\mu \in \mathcal{P}(\mathbb{R}^N)$ we have $\alpha_{\mathcal{B}_{\mathcal{K}}}(\mu_n(\omega), \mu) \xrightarrow{n\to\infty} 0$ for P-almost all $\omega \in \Omega$, thus obtaining the following as a consequence of Theorem 3.2.5.

Theorem 3.2.8. *Let the technology matrix be fixed and assume (B1)-(B2). Furthermore, for some $p > 1$ and $K > 0$ let $\mu \in \mathcal{P}_{p,K}(\mathbb{R}^{m_2})$, and let $V \subseteq \mathbb{R}^{n_1}$ be a bounded open set such that $\Psi_V(\mu)$ is a CLM set for $TSMIR(\mu)$ with respect to V. Then,*

(a) $\varphi_V(\mu_n(\omega)) \xrightarrow{n\to\infty} \varphi_V(\mu)$ for P-almost all $\omega \in \Omega$;

(b) for any open set $G \subseteq \mathbb{R}^{n_1}$ with $\Psi_V(\mu) \subseteq G$ and for P-almost all $\omega \in \Omega$ there exists some $n_0(\omega) \in \mathbb{N}$ such that $\Psi_V(\mu_n(\omega)) \subseteq G$ for all $n \geq n_0(\omega)$;

(c) for P-almost all $\omega \in \Omega$ there exists some $n_1(\omega) \in \mathbb{N}$ such that $\Psi_V(\mu_n(\omega))$ is a CLM set for $TSMIR(\mu_n(\omega))$ with respect to V for all $n \geq n_1(\omega)$.

3.3 Solution Procedures

Two-stage stochastic integer programming problems have proved particularly difficult to solve, since they suffer from the combined hardships of stochastic programming and integer programming. In this section we give a brief account of some of the general purpose algorithms that have been proposed for this class of problems, but one should be aware that specialized solution procedures will be required to solve many practical problems cf. also Chapters 6 through 10 of this thesis. Let us also note that very few attempts have been made to elaborate general purpose algorithms for multistage stochastic integer programs, and in fact only few results on structural properties are available for this class of problems (see e.g. Römisch and Schultz [129]). As in the linear recourse case, most solution procedures for two-stage stochastic programs with (mixed-) integer recourse are based on the following assumption.

(B3) The distribution μ of ξ is discrete and has finite support, say $\Xi = \{\xi^1, \ldots, \xi^S\}$ with corresponding probabilities $p^1, \ldots, p^S$.

Again, for each $s \in \{1, \ldots, S\}$ the outcome of random parameters $\big(h(\xi^s), T(\xi^s)\big)$, corresponding to some elementary event $\xi^s \in \Xi$, is referred to as a *scenario*, and we denote it simply by (h^s, T^s).

Remark 3.3.1. Cf. Remark 2.3.1 on page 23, we note that (B3) may be justified by the stability results presented in the previous section. In particular, according to Theorem 3.2.6 the optimal value and the solution set of a stochastic program with an absolutely continuous distribution of random parameters, may be approximated to any given accuracy by the optimal value and the solution set of problems employing only discrete distributions.

As in the linear recourse case, most solution procedures for stochastic programs with (mixed-) integer recourse are based on decomposition of the problem, and in general they can be classified as either *primal decomposition procedures* in which the problem is decomposed with respect to the individual stages, or *dual decomposition procedures* in which the problem is decomposed with respect to the individual outcomes of random parameters (see e.g. Carøe [27]). A notable exception is the approach of Hemmecke and Schultz [55], where a two-stage stochastic program with integer recourse and random second-stage right-hand side is solved as a large-scale integer program by an augmentation algorithm using its Graver test set (see Graver [51]), and decomposition is applied not to the problem itself but to the Graver test set.

In Sections 3.3.1 and 3.3.2 below we discuss in some detail a few primal and dual decomposition procedures, respectively. Before proceeding to this, let us mention that van der Vlerk [158] proposed a convex approximation of the expected recourse function for problems with complete integer recourse and random second-stage right-hand side, and showed that the approximation provides a convex lower bound that is strictly better than the one provided by the LP-relaxation (and in fact optimal for problems with a totally unimodular recourse matrix). We also mention that Carøe and Tind [30] discussed the use of cutting plane techniques for two-stage mixed 0-1 recourse models in relation to primal as well as dual decomposition procedures. Finally, we note that the stochastic branch-and-bound procedure discussed in general settings by Norkin, Ermoliev, and Ruszczyński [98] and Norkin, Pflug, and Ruszczyński [99], may be applied, in particular, to the two-stage (mixed-) integer recourse model considered here.

3.3.1 Primal Decomposition

When integer variables occur only in the first stage the difficulties are not too severe, since the structural properties of the expected recourse function discussed in Chapter 2 are maintained. Hence in this case it is relatively easy to adapt the L-shaped algorithm presented in Section 2.3.1 to account for integer first-stage variables. Such an approach was first formalized by Wollmer [166] for problems with binary first-stage variables and continuous second-stage variables (see also Chapters 7 and 10 of this thesis). Generalizations of this approach for problems with (mixed-) integer recourse have been suggested by various authors. In general these procedures are based on the following equivalent formulation of the stochastic program (3.1.1),

$$\min\ cx + \theta, \tag{3.3.1a}$$
$$\text{s.t. } Ax = b, \tag{3.3.1b}$$
$$\theta \geq \mathcal{Q}(x, \mu), \tag{3.3.1c}$$
$$x \in K_2, \tag{3.3.1d}$$
$$x \in X,\ \theta \in \mathbb{R}, \tag{3.3.1e}$$

where $K_2 = \{x \in \mathbb{R}_+^{n_1} \mid \mathcal{Q}(x, \mu) < \infty\}$. We recall that the restriction $x \in K_2$ is referred to as the *induced constraints*, and note that under assumption (B3) it implies that the first-stage decision must be such that the second-stage problem is feasible for all scenarios. Now, as in L-shaped decomposition for two-stage linear recourse problems, the fundamental idea is to relax the constraints (3.3.1c) and (3.3.1d), and iteratively re-enforce them by imposing so-called optimality cuts and feasibility cuts, respectively. Furthermore, this approach must now be combined with a branching strategy to account for integer requirements on first-stage variables. Hence, Laporte and Louveaux [82] suggested a conceptual integer L-shaped algorithm, formulated as a branch-and-cut algorithm as follows.

Algorithm 3.1 *(Integer L-shaped Decomposition)*

Step 1 (*Initialization*) Let $K > 0$, set $\bar{z} = \infty$, and let the list of open problems $\mathcal{L}$ consist of the problem $\min\{cx + \theta \mid Ax = b,\ x \in \mathbb{R}_+^{n_1},\ \theta \in \mathbb{R}\}$.

Step 2 (*Termination/Node selection*) If $\mathcal{L} = \emptyset$, stop; the solution that yielded the upper bound $\bar{z}$ is optimal. Otherwise, select and remove a problem P from $\mathcal{L}$.

Step 3 (*Solve master problem*) Solve the current problem and let (x^P, θ^P) be an optimal solution vector if one exists; if the problem is unbounded, let (x^P, θ^P) be a feasible solution with $cx^P + \theta^P < \bar{z} - K$. If $cx^P + \theta^P \geq \bar{z}$, go to Step 2.

Step 4 (*Add feasibility cuts*) Check for any induced constraint violations. If one exists, add a feasibility cut to the current problem and return to Step 3.

Step 5 (*Branching*) Check for integrality restrictions. If a restriction is violated by the current solution x^P, create two new problems by branching on the relevant component of x, add the two problems to $\mathcal{L}$, and return to Step 2.

Step 6 (*Update bound*) Compute $\mathcal{Q}(x^P, \mu)$ and let $\bar{z} = \min\{\bar{z},\ cx^P + \mathcal{Q}(x^P, \mu)\}$.

Step 7 (*Add optimality cuts*) If $\theta^P < \mathcal{Q}(x^P, \mu)$, add an optimality cut to the current problem and return to Step 3. Otherwise, go to Step 2.

Clearly the convergence of this algorithm to an optimal solution of problem (3.1.1) requires the existence of sets of optimality cuts and feasibility cuts that are sufficient to re-enforce the constraints (3.3.1c) and (3.3.1d), respectively. Laporte and Louveaux [82] considered a two-stage stochastic program with complete (mixed-) integer recourse, binary first-stage variables, and general but easily computable recourse problems. The authors proposed optimality cuts that approximate the expected recourse function at binary first-stage solutions (but not necessarily at other points), obtaining finite convergence by virtue of finiteness of the set of feasible first-stage solutions. This algorithm has been successfully applied to several problems cf. e.g. Laporte, Louveaux, and Mercure [83, 84] and Laporte, Louveaux, and van Hamme [85]. In a more general setting, Carøe and Tind [31] proposed an L-shaped method for two-stage integer recourse problems, based on duality theory for integer programming (see e.g. Nemhauser and Wolsey [97]). Deriving feasibility cuts and optimality cuts defined by non-linear dual price functions, the authors establish finite convergence of the algorithm, but admit that no practicable method for the solution of the resulting master problem exists.

Another primal decomposition approach was suggested by Schultz, Stougie, and van der Vlerk [143] for problems with continuous first stage and a second-stage value function given by $\Phi(\tau) = \min\{qy \mid Wy \geq \tau,\ y \in \mathbb{Z}_+^{n_2}\}$. The following assumptions are employed.

(B4) The technology matrix is fixed, i.e. $T^s = T$ for $s = 1, \ldots, S$.

(B5) The recourse matrix W is integral.

Now, the insight of Schultz et al. was to observe that, for all $\bar{x} \in \mathbb{R}^{n_1}$, the expected recourse function $\mathcal{Q}$ is constant on the set

$$C(\bar{x}) = \bigcap_{s=1}^{S} \{x \in \mathbb{R}^{n_1} \mid \lceil Tx - h^s \rceil = \lceil T\bar{x} - h^s \rceil\},$$

where $\lceil \cdot \rceil$ denotes componentwise integer round up. Using the fact that the expected value function is lower semicontinuous (cf. Theorem 3.2.1), the authors now show that $V = \{x \in \mathbb{R}^{n_1} \mid x \text{ is a vertex of } X \cap C(x)\}$ is a countable set containing an optimal solution of the problem. Furthermore, the authors show how level sets of the linear relaxation of the problem can be used to reduce V to a finite set, still containing an optimal solution. Finally, an enumeration scheme, taking advantage of the structure of the set V, is suggested. The method requires a potentially large number of function evaluations of the expected recourse function, each of which requires S second-stage problems to be solved. To this end, the authors propose to exploit the similarity of the second-stage problems, solving them by means of Gröbner basis methods. Clearly, though, the algorithm is independent of the particular way that function evaluations are performed.

The method of Schultz et al. was extended by Ahmed, Tawarmalani, and Sahinidis [1] as follows. First of all note, however, that the problem under consideration is stated slightly different than problem (3.1.1), as we are now concerned with the following,

$$\min\left\{cx + \sum_{s=1}^{S} p^s q^s y^s \;\middle|\; -Tx + W^s y^s \geq h^s,\ x \in X,\ y^s \in \mathbb{Z}_+^{n_2},\ s = 1, \ldots, S\right\}.$$

Introducing the variable transformation $\chi = Tx$, Ahmed et al. propose the following reformulation of this problem,

$$\min_{\chi \in \mathcal{X}} \left\{F(\chi) = f(\chi) + \overline{\Psi}(\chi)\right\} \tag{3.3.2}$$

where

$$f(\chi) = \min\{cx \mid Tx = \chi,\ x \in X\},$$

$$\overline{\Psi}(\chi) = \sum_{s=1}^{S} p^s \min\{q^s y \mid W^s y \geq h^s + \chi,\ y \in \mathbb{Z}_+^{n_2}\},$$

and

$$\mathcal{X} = \{\chi \in \mathbb{R}^{m_2} \mid \exists\, x \in X : Tx = \chi\}.$$

The transformed problem (3.3.2) is easily seen to be equivalent to the original problem in the sense that if $\chi^* \in \mathcal{X}$ is an optimal solution of the transformed problem, then any $x^* \in \arg\min\{cx \mid Tx = \chi^*,\ x \in X\}$ is an optimal solution of the original problem, and the optimal objective values are identical cf. [1, Theorem 3.2]. Now, Ahmed et al. extended the above-mentioned observation of Schultz et al. to note that for any $k \in \mathbb{Z}^{m_2 S}$ the recourse function $\overline{\Psi}$ is constant on

$$\mathcal{C}(k) = \bigcap_{s=1}^{S} \prod_{j=1}^{m_2} (k_j^s - h_j^s - 1, k_j^s - h_j^s].$$

Assuming that the feasible set $\mathcal{X}$ is non-empty and compact, the authors then propose a branch-and-bound procedure, where branching occurs along the possible discontinuities of $\overline{\Psi}$. In particular, the algorithm proceeds by partitioning $\mathcal{X}$ into regions of the form $\mathcal{X} \cap \Pi_{j=1}^{m_2}(l_j, u_j]$, where each l_j, $j = 1, \ldots, m_2$, is such that $l_j + h_j^s$ is integral for some $s \in \{1, \ldots, S\}$. This is combined with a specialized bounding procedure as follows.

Algorithm 3.2 *(Ahmed et al. [1])*

Step 1 (*Initialization*) Set $\bar{z} = \infty$. Let $l^P, u^P \in \mathbb{R}^{m_2}$ be such that $\mathcal{X} \subseteq \Pi_{j=1}^{m_2}\,(l_j^P, u_j^P]$ and for all $j = 1, \ldots, m_2$, $l_j^P + h_j^s$ is integral for some $s \in \{1, \ldots, S\}$. Let the list of open problems $\mathcal{L}$ consist of problem P defined by (3.3.2) with the additional constraints $l^P < \chi \leq u^P$. Also, let $\epsilon \in \mathbb{R}_+^{m_2}$ be such that $\overline{\Psi}(\cdot)$ is constant over $\Pi_{j=1}^{m_2}\,(l_j, l_j + \epsilon_j]$ whenever $l \in \mathbb{R}^{m_2}$ is such that for all $j = 1, \ldots, m_2$, $l_j + h_j^s$ is integral for some $s \in \{1, \ldots, S\}$.

Step 2 (*Termination/Node selection*) If $\mathcal{L} = \emptyset$, stop; the solution that yielded the upper bound $\bar{z}$ is optimal. Otherwise, select and remove from $\mathcal{L}$ a problem P, defined as $\inf\{F(\chi) \mid l^P < \chi \leq u^P,\ \chi \in \mathcal{X}\}$.

Step 3 (*Bounding*) Obtain a lower bound on P by solving the lower bounding problem $z^P = \overline{\Psi}(l^P + \epsilon) + \min\{cx \mid Tx = \chi,\ l^P \leq \chi \leq u^P,\ x \in X\}$ and let χ^P be an optimal solution. If $z^P \geq \bar{z}$ go to Step 2. Otherwise, let $\bar{z} = \min\{\bar{z}, F(\chi^P)\}$ and remove from $\mathcal{L}$ all problems P' with $z^{P'} \geq \bar{z}$.

Step 4 (*Branching*) Select an index $j \in \{1, \ldots, m_2\}$ and a value v_j such that $v_j + h_j^s$ is integral for some $s \in \{1, \ldots, S\}$ and $l_j^P < v_j < u_j^P$. Construct two new problems P' and P'', obtained from P by adding the constraints $\chi_j > v_j$ and $\chi_j \leq v_j$, respectively. Let $z^{P'} = z^{P''} = z^P$ and add the two problems to $\mathcal{L}$. Go to Step 2.

Let us note that Ahmed et al. presented a procedure for the a priori determination of the constant ϵ, simply determining the smallest possible width of the non-empty regions $C(k)$, $k \in \mathbb{Z}^{m_2 S}$. The authors proved finite convergence of the algorithm, exploiting the fact that only a finite number of non-empty regions $C(k) \cap \mathcal{X}$, $k \in \mathbb{Z}^{m_2 S}$, exist when $\mathcal{X}$ is compact. Finally, enhancements of the lower bounding procedure and extension of the algorithm to the case of random technology matrix are discussed.

3.3.2 Dual Decomposition

In this section we consider in particular the dual decomposition procedure proposed by Carøe and Schultz [29] for the two-stage stochastic program with mixed-integer recourse defined by (3.1.1). (In fact the authors allow also for a random recourse matrix.) The fundamental idea is to use the variable splitting approach, first proposed for combinatorial optimization problems by Jörnsten, Näsberg, and Smeds [64], introducing copies of the first-stage variables $x^1, \dots, x^S$ for each scenario, and writing the non-anticipativity constraint $x^1 = \cdots = x^S$ explicitly as $\sum_{s=1}^{S} H^s x^s = 0$, where $H = (H^1, \dots, H^S)$ is a suitably defined matrix of size $l \times n_1 S$. (See Carøe [27] for a discussion of possible alternative formulations of the non-anticipativity constraints.) Hence the problem is formulated as

$$\begin{aligned} \min\ & \sum_{s=1}^{S} p^s (c x^s + q^s y^s) \\ \text{s.t. } & A x^s = b, && s = 1, \dots, S, \\ & T^s x^s + W^s y^s = h^s, && s = 1, \dots, S, \\ & \sum_{s=1}^{S} H^s x^s = 0, \\ & x^s \in X,\ y^s \in Y, && s = 1, \dots, S. \end{aligned} \tag{3.3.3}$$

Using a Lagrangian relaxation of the non-anticipativity constraints, the authors obtain a problem that is separable into independent scenario subproblems, defined for Lagrange multipliers $\lambda \in \mathbb{R}^l$ by

$$D(\lambda) = \sum_{s=1}^{S} \min\big\{ L^s(x^s, y^s, \lambda) \mid A x^s = b,\ T^s x^s + W^s y^s = h^s,\ x^s \in X,\ y^s \in Y \big\}.$$

where $L^s(x^s, y^s, \lambda) = p^s(c x^s + q^s y^s) + \lambda(H^s x^s)$ for $s = 1, \dots, S$. Now, by a well-known weak duality result (see e.g. Nemhauser and Wolsey [97]), a lower bound on the optimal value of (3.3.3) is obtained by solving the Lagrangian dual,

$$z_{LD} = \max\big\{ D(\lambda) \mid \lambda \in \mathbb{R}^l \big\}.$$

Moreover, if for some choice of Lagrange multipliers λ, the corresponding solutions (x^s, y^s), $s = 1, \dots, S$, of the Lagrangian relaxation $D(\lambda)$, satisfy the non-anticipativity constraints, then (x^s, y^s), $s = 1, \dots, S$, is an optimal solution of (3.3.3), and λ is an optimal solution of the Lagrangian dual. In general, though, a duality gap persists, and the approach must be combined with a branching scheme to enforce the non-anticipativity constraints. This is formalized in the following algorithm.

Algorithm 3.3 *(Dual Decomposition)*

Step 1 (*Initialization*) Set $\bar{z} = \infty$ and let the list of open problems $\mathcal{L}$ consist of problem P defined by (3.3.3).

Step 2 (*Termination/Node selection*) If $\mathcal{L} = \emptyset$, stop; the solution that yielded the upper bound $\bar{z}$ is optimal. Otherwise, select and remove a problem P from $\mathcal{L}$.

Step 3 (*Bounding*) Solve the Lagrangian dual of P to obtain the lower bound z_{LD}^P. If $z_{LD}^P \geq \bar{z}$, return to Step 2. Otherwise,

(i) if $x^s = \bar{x}$ for all $s \in \{1, \dots, S\}$, let $\bar{z} = \min\{\bar{z}, c\bar{x} + \mathcal{Q}(\bar{x}, \mu)\}$, remove from $\mathcal{L}$ all problems P' with $z_{LD}^{P'} \geq \bar{z}$, and return to Step 2;

(ii) if $x^1, \dots, x^S$ differ, compute the weighted average $\bar{x} = \sum_{s=1}^S p^s x^s$ and round it by some suitable heuristic to obtain a first-stage solution $\bar{x}^R$. If $\bar{x}^R$ is feasible, let $\bar{z} = \min\{\bar{z}, c\bar{x}^R + \mathcal{Q}(\bar{x}^R, \mu)\}$, and remove from $\mathcal{L}$ all problems P' with $z_{LD}^{P'} \geq \bar{z}$.

Step 4 (*Branching*) Select an index $i \in \{1, \dots, n_1\}$ such that $x_i^1, \dots, x_i^S$ differ, and construct two new problems P' and P'', obtained from P by adding the constraints $x_i \leq \lfloor \bar{x}_i \rfloor$ and $x_i \geq \lceil \bar{x}_i \rceil$, respectively. Let $z_{LD}^{P'} = z_{LD}^{P''} = z_{LD}^P$ and add the two problems to $\mathcal{L}$. Go to Step 2.

The rounding heuristic in Step 3 (ii) is employed to more quickly generate feasible solutions improving the initial upper bound, thereby possibly speeding up the procedure. The algorithm has been successfully implemented and was tested on a realistic two-stage unit commitment problem in Carøe, Ruszczyński, and Schultz [28] (see also Carøe [27]). As in the linear recourse case, the dual decomposition approach is applicable, in theory, also to multistage problems. From an implementational viewpoint, however, some work remains to be done, since the dimension of the Lagrange multiplier grows drastically in this case.

Finally, let us note that Løkketangen and Woodruff [87] and Takriti, Birge, and Long [153] applied the progressive hedging algorithm to multistage stochastic mixed-integer programming problems. Even though this method is only formally justified for linear recourse problems, the authors observe convergence of the algorithm, obtaining good quality solutions.

Part II

Non-Classical Stochastic Programs

Chapter 4

The Minimum Risk Problem

In Chapter 1 we introduced the classical two-stage stochastic program with recourse, and in Chapters 2 and 3 we discussed this class of problems in more detail, considering the linear recourse case and the mixed-integer recourse case, respectively. We recall that the problem is to determine in a first stage a here-and-now decision that must be made without complete knowledge on some uncertain parameters of the model. This first-stage decision must be made so as to minimize the sum of direct cost and the expected value of future recourse cost. The recourse cost, on the other hand, is determined in a second stage as some recourse decisions can be made after uncertainty has been revealed. Several objections may be put forward against the formulation of this classical stochastic program, a primary objection being that minimization of the expected cost does not always constitute an appropriate objective. The appropriateness of this criterion is dependent on the assumption that the decision process is to be repeated a great number of times, implying by the law of large numbers that, in the long run, average cost will be equal to the expected cost. This assumption, however, will frequently not be justified and consequently the expected cost may not be of much interest to the decision-maker. Another major objection against the expected cost as the object of minimization, is the fact that the optimal solution of the classical stochastic program may only assure the achievement of the corresponding minimum expected cost with a relatively small probability. These considerations suggest that the risk averse decision-maker may not consider the solution of the classical stochastic program to be "optimal". As an alternative, we consider here the problem of minimizing the probability of total cost exceeding some prescribed threshold value ϕ, that may be thought of as the level of bankruptcy or even just a budget limit. This approach may in fact be seen as an application of the minimum risk criterion (see e.g. Bereanu [13]) in the two-stage recourse setting described above, and hence we refer to the resulting formulation as the *minimum risk problem*.

4.1 Problem Formulation

In this chapter we consider the minimum risk problem, assuming that the second-stage problem may be appropriately modeled as a linear programming problem. For generalization of the results to the case of mixed-integer recourse, we refer to Schultz [137] and Schultz and Tiedemann [144]. (See also the masters thesis by Tiedemann [154].) During the structural analysis of classical two-stage stochastic programs with mixed-integer recourse, presented in Chapter 3, we assumed that the recourse matrix and the second-stage cost are fixed, and this assumption is maintained here. Hence we let $\xi : \Omega \mapsto \mathbb{R}^N$ be a random vector defined on some probability space $(\Omega, \mathcal{F}, P)$, the components of which constitute the random second-stage data, consisting of the second-stage right-hand side $\tilde{h} : \Omega \mapsto \mathbb{R}^{m_2}$ and the technology matrix $\tilde{T} : \Omega \mapsto \mathbb{R}^{m_2 \times n_1}$. In other words we have $N = m_2(1 + n_1)$, and for $\omega \in \Omega$ we have $\xi(\omega) = \big(\tilde{h}(\omega), \tilde{T}_1(\omega), \ldots, \tilde{T}_{m_2}(\omega)\big)$ where $\tilde{T}_k$ denotes the kth row of $\tilde{T}$ for $k = 1, \ldots, m_2$. Introducing the induced probability measure $\mu = P \circ \xi^{-1}$ on $\mathbb{R}^N$, the minimum risk problem is now formally stated as,

$$MRP(\mu) \qquad \min\big\{\mathcal{Q}_P(x, \mu) \mid Ax = b,\ x \in X\big\}, \tag{4.1.1}$$

where the recourse function $\mathcal{Q}_P(x, \mu)$ denotes the probability of total cost exceeding the threshold value ϕ, given the first-stage decision x,

$$\mathcal{Q}_P(x, \mu) = \mu\big(\big\{\xi \in \mathbb{R}^N \mid cx + \Phi\big(h(\xi) - T(\xi)x\big) > \phi\big\}\big), \tag{4.1.2}$$

and the second-stage value function Φ is defined by

$$\Phi(\tau) = \min\big\{qy \mid Wy = \tau,\ y \in \mathbb{R}^{n_2}_+\big\}. \tag{4.1.3}$$

Here we assume that $c \in \mathbb{R}^{n_1}$, $q \in \mathbb{R}^{n_2}$, and $b \in \mathbb{R}^{m_1}$ are known vectors, and that $A \in \mathbb{R}^{m_1 \times n_1}$ and $W \in \mathbb{R}^{m_2 \times n_2}$ are known matrices. The second-stage right-hand side $h : \mathbb{R}^N \mapsto \mathbb{R}^{m_2}$ and the technology matrix $T : \mathbb{R}^N \mapsto \mathbb{R}^{m_2 \times n_1}$, on the other hand, are represented as mappings picking out the appropriate components of the random vector ξ. The set $X \subseteq \mathbb{R}^{n_1}_+$ is assumed to be non-empty and closed, and in particular it may or may not contain integrality restrictions on some or all of the first-stage variables.

In the following we let $\mu \in \mathcal{P}(\mathbb{R}^N)$ represent the joint distribution of the second-stage right-hand side $\tilde{h}$ and the technology matrix $\tilde{T}$ as described above. At some points we will also be interested in the marginal distributions of $\tilde{h}$ and $\tilde{T}$, and we denote these by μ_1 and μ_2, respectively. Moreover, for $T \in \mathbb{R}^{m_2 \times n_1}$ we denote by $\mu_1^2(\cdot, T)$ the (regular) conditional distribution of $\tilde{h}$ given $\tilde{T} = T$. (See Appendix A.)

4.1.1 Mean-Risk Models

Note that the formulation of the minimum risk problem (4.1.1), considered throughout this chapter, will typically not be the one applied in practice, since the decision-maker will most likely find it more appropriate to incorporate the probability-based recourse function (4.1.2) into a bicriterion *mean-risk model*,

$$\begin{aligned} \min\ & cx + \mathcal{Q}_E(x, \mu), \\ \min\ & \mathcal{Q}_P(x, \mu), \\ \text{s.t. } & Ax = b, \\ & x \in X, \end{aligned} \tag{4.1.4}$$

where $\mathcal{Q}_E(x, \mu) = \mathbb{E}\big[\Phi\big(h(\xi) - T(\xi)x\big)\big]$ is the classical expected recourse function considered in Chapter 2. The mean-risk model (4.1.4) allows a simple tradeoff analysis, either analytical or geometrical, between the expected cost and the associated risk, the latter expressed here as the probability of total cost exceeding the threshold value. In particular, a common approach to this bicriterion problem is to solve single-objective mean-risk problems of the form

$$\min\big\{cx + \mathcal{Q}_E(x,\mu) + \alpha\mathcal{Q}_P(x,\mu) \mid Ax = b,\ x \in X\big\}, \tag{4.1.5}$$

where $\alpha > 0$ is a weight factor.

Remark 4.1.1. To ease the exposition we find it more convenient throughout this chapter to consider the minimum risk problem in its self-contained form (4.1.1) rather than the perhaps more appropriate mean-risk model (4.1.4). It is straightforward, though, to extend the structural results presented in the subsequent sections for the minimum risk problem, to apply also to mean-risk problems of the form (4.1.5), simply by complementing the assumptions made here with those employed during the survey of structural results for the classical stochastic program in Chapter 2. Also, the solution procedure presented in Section 4.3 for the minimum risk problem closely resembles the L-shaped algorithm discussed in Chapter 2, and hence a hybrid version of the two algorithms for problem (4.1.5) is only natural cf. also our discussion in Section 4.3.1 below. In Chapter 8 we consider an application of the bicriterion mean-risk model (4.1.4) in more detail.

The mean-risk problem (4.1.4) may in general be understood as the application of two particular scalar criteria for the selection of a "best" random variable from the collection $\{C_x\}_{x\in K_1}$, where $C_x = cx + \Phi\big(h(\xi) - T(\xi)x\big)$ represents total (random) cost arising from the first-stage decision $x \in K_1 = \{x \in X \mid Ax = b\}$. A more general approach to this problem of choosing among uncertain outcomes is provided by the relation of *stochastic dominance* that introduces a partial order in the space of real random variables (see e.g. Whitmore and Findlay [165]). Stochastic dominance is theoretically attractive, but provides severe computational difficulties as a multiobjective model with a continuum of objectives. Therefore, it is of some importance to establish relationships between a particular mean-risk model and the relation of stochastic dominance, guaranteeing that the solution of the mean-risk model is not stochastically dominated by other feasible solutions. In particular, in our cost minimization framework, smaller outcomes are preferred to larger, and hence we define first degree stochastic dominance in the following way.

Definition 4.1.1. Let $\{C_x\}_{x\in K}$ be a family of random variables defined on some probability space $(\Omega, \mathcal{F}, P)$. For $x', x'' \in K$ we say that x' stochastically dominates x'' to first degree if $P(C_{x'} \le \eta) \ge P(C_{x''} \le \eta)$ for all $\eta \in \mathbb{R}$, and in that case we write $x' \succeq x''$.

Considering the family of stochastic variables $\{C_x\}_{x\in K_1}$ representing total random cost as defined above, i.e. $C_x = cx + \Phi\big(h(\xi) - T(\xi)x\big)$ such that $\mathbb{E}[C_x] = cx + \mathcal{Q}_E(x,\mu)$ and $P(C_x > \phi) = \mathcal{Q}_P(x,\mu)$ for all $x \in K_1$, the following result is obvious.

Lemma 4.1.1. *The mean-risk problem (4.1.4) is consistent with first degree stochastic dominance, meaning that for $x', x'' \in K_1$ we have*

$$x' \succeq x'' \quad \Rightarrow \quad cx' + \mathcal{Q}_E(x',\mu) \le cx'' + \mathcal{Q}_E(x'',\mu) \quad \textit{and} \quad \mathcal{Q}_P(x',\mu) \le \mathcal{Q}_P(x'',\mu)$$

Furthermore, as an immediate consequence we have the following result.

Lemma 4.1.2. *For any $\alpha > 0$ the mean-risk problem (4.1.4) is α-consistent with first degree stochastic dominance, meaning that for $x', x'' \in K_1$ we have*

$$x' \succeq x'' \quad \Rightarrow \quad cx' + \mathcal{Q}_E(x', \mu) + \alpha \mathcal{Q}_P(x', \mu) \leq cx'' + \mathcal{Q}_E(x'', \mu) + \alpha \mathcal{Q}_P(x'', \mu)$$

Remark 4.1.2. The result in Lemma 4.1.2 was also pointed out by Schultz and Tiedemann [144], who considered the minimum risk problem in the mixed-integer recourse case. It provides a direct connection between stochastic dominance and mean-risk problems of the form (4.1.5), in the sense that an optimal solution of problem (4.1.5) cannot be stochastically dominated to first degree by any non-optimal feasible solution

Remark 4.1.3. In this chapter we only consider the risk measure $\mathcal{Q}_P$ defined by (4.1.2). Clearly, though, alternative measures of risk are possible. The selection of an appropriate measure of risk is dependent not only on the theoretical issue of consistencies with stochastic dominance as discussed above, but also on issues such as smoothness and convexity properties, as well as applicability of practicable solution procedures for resulting mean-risk problems such as (4.1.5). In this respect, we have seen that the risk measure $\mathcal{Q}_P$ leads to a mean-risk model that is consistent with first degree stochastic dominance, and in the subsequent sections we will establish structural properties, leading to the elaboration of a practicable solution procedure for problem (4.1.5). For a discussion of other risk measures, satisfying stronger consistencies with stochastic dominance, we refer to Ogryczak and Ruszczyński [100, 101, 102, 103]. See also Artzner et al. [5] where a class of so-called *coherent measures of risk* was defined by means of a number of axioms.

4.2 Structural Properties

In this section we will show that under reasonably mild assumptions the minimum risk problem (4.1.1) is equivalent to a classical stochastic program in which a binary variable and an additional constraint are included in the second stage. Thus, the minimum risk problem belongs to the class of classical two-stage stochastic programs with mixed-integer recourse. Obviously, though, the structure of the minimum risk problem is much simpler than that of a general stochastic program with mixed-integer recourse, and hence we will show that some of the results discussed in Chapter 3, remain valid for the minimum risk problem under less restrictive assumptions. Early results in this direction were obtained by Raik [110, 111] who established lower semicontinuity of the recourse function $\mathcal{Q}_P(\cdot, \mu)$, as well as a sufficient condition for continuity. (See also Kibzun and Kan [70].) Lower semicontinuity of $\mathcal{Q}_P(\cdot, \mu)$ can also be derived from Proposition 3.1 in Römisch and Schultz [126], a statement concerning chance-constrained stochastic programs.

Throughout the structural analysis we will make just the following two assumptions, ensuring that the second-stage value function Φ is real-valued.

(C1) For all $t \in \mathbb{R}^{m_2}$ there exists $y \in \mathbb{R}^{n_2}_+$ such that $Wy = t$.

(C2) There exists $u \in \mathbb{R}^{m_2}$ such that $uW \leq q$.

Here (C1) is the assumption of complete recourse, ensuring feasibility of the second-stage problem for all possible right-hand sides, while (C2) is employed to ensure dual feasibility and hence boundedness of the second-stage problem. From Lemma 2.2.1 on page 17 we recall that under assumptions (C1) and (C2), Φ is a real-valued, piecewise linear, and convex function on $\mathbb{R}^{m_2}$.

Remark 4.2.1. As pointed out also in Chapter 2, we note that for practical purposes it is often sufficient to replace (C1) by the weaker assumption of relatively complete recourse, ensuring feasibility of the second-stage problem only for those right-hand sides that correspond to a feasible first-stage solution and a possible outcome of random parameters. Hence, denoting by $\Xi \subseteq \mathbb{R}^N$ the support of ξ, i.e. the smallest closed subset such that $\mu(\Xi) = 1$, we may assume that for all $x \in X$ satisfying $Ax = b$ and for all $\xi \in \Xi$ there exists $y \in \mathbb{R}^{n_2}_+$ such that $Wy = h(\xi) - T(\xi)x$ (cf. Definition 2.2.2 on page 17). Furthermore, we note that if relatively complete recourse is not inherent in the problem, it may be established by the inclusion of feasibility cuts cf. for example the discussion of the L-shaped algorithm in Section 2.3.1.

4.2.1 The Equivalent Classical Stochastic Program

Evidently, the objective of the minimum risk problem (4.1.1) may be equivalently formulated as the expected value of an appropriately defined indicator function. Specifically, for $x \in \mathbb{R}^{n_1}$ we may define the set of all outcomes of random parameters yielding total cost exceeding the threshold value,

$$M(x) = \big\{\xi \in \mathbb{R}^N \mid cx + \Phi\big(h(\xi) - T(\xi)x\big) > \phi\big\} \tag{4.2.1}$$

and introduce the corresponding indicator function, $\psi : \mathbb{R}^{n_1} \times \mathbb{R}^N \mapsto \{0, 1\}$,

$$\psi(x,\xi) = \begin{cases} 1 & \text{if} \quad \xi \in M(x); \\ 0 & \text{otherwise.} \end{cases} \tag{4.2.2}$$

We now have

$$\mathcal{Q}_P(x,\mu) = \mu\big(M(x)\big) = \int_{\mathbb{R}^N} \psi(x,\xi)\mu(d\xi). \tag{4.2.3}$$

In some situations, it is possible to define the indicator function (4.2.2) as the value function of a mixed-integer program, i.e.

$$\begin{aligned} \psi(x,\xi) = \min\big\{\theta \mid Wy = h(\xi) - T(\xi)x,& \\ qy - M_0\theta \le \phi - cx,\ y \in \mathbb{R}^{n_2}_+,\ \theta \in \{0,1\}\big\},& \end{aligned} \tag{4.2.4}$$

where $M_0 > 0$ is some “sufficiently large” number. In particular, denoting again by $\Xi \subseteq \mathbb{R}^N$ the support of ξ and observing (C1), we may note that a sufficient condition for (4.2.4) to be feasible for all $x \in \big\{x \in X \mid Ax = b\big\}$ and $\xi \in \Xi$, is to have

$$M_0 = \sup\big\{cx + \Phi\big(h(\xi) - T(\xi)x\big) - \phi \;\big|\; Ax = b,\ x \in X,\ \xi \in \Xi\big\}. \tag{4.2.5}$$

Clearly, if $\big\{x \in X \mid Ax = b\big\}$ and Ξ are both bounded, the supremum exists and is finite. In fact, Schultz and Tiedemann [144] considered the minimum risk problem in the mixed-integer recourse case, and gave a formal proof, showing that if $\big\{x \in X \mid Ax = b\big\}$ and

Ξ are both bounded, then the minimum risk problem (4.1.1) with the recourse function $\mathcal{Q}_P$ defined by either (4.1.2)-(4.1.3) or (4.2.3)-(4.2.5), respectively, are equivalent. Apart from showing that (4.2.4) does in fact provide an appropriate definition of the indicator function (4.2.2) when M_0 is defined by (4.2.5), this proof takes care of the measurability of ψ, and hence shows that (4.2.3) is well-defined.

Hence we see that under mild assumptions, the minimum risk problem (4.1.1)-(4.1.3) is in fact equivalent to a classical two-stage stochastic program with mixed-integer recourse. As pointed out by Schultz and Tiedemann [144], though, the equivalent classical stochastic program has relatively complete mixed-integer recourse, but fails to satisfy the complete mixed-integer recourse property. Thus, the continuity properties of the expected recourse function discussed in Chapter 3, cannot be directly applied to the probability-based recourse function $\mathcal{Q}_P$. Still, as mentioned above, the special structure of the minimum risk problem in fact causes some of the results established in Chapter 3 to remain valid for the minimum risk problem even under less restrictive assumptions. In particular, the assumption of boundedness of $\big\{x \in X \mid Ax = b\big\}$ and Ξ, required for the existence of an equivalent classical stochastic program, turns out to be of no importance for the structural analysis of the minimum risk problem, and moreover, the results are maintained even without the moment conditions employed in Chapter 3.

4.2.2 The Probability-Based Recourse Function

In Section 4.2.1 above we defined for all $x \in \mathbb{R}^{n_1}$ the set $M(x)$ of all those outcomes of random parameters that yield total cost exceeding the threshold value cf. (4.2.1). When studying the structural properties of $\mathcal{Q}_P$ as a function of x, we will also find it convenient to define for $x \in \mathbb{R}^{n_1}$ the set $E(x)$ of all those $\xi \in \mathbb{R}^N$ such that the indicator function $\psi(\cdot, \xi)$, defined by (4.2.2), is discontinuous at x. By continuity of Φ this set is easily seen to be equal to the set of all those outcomes of random parameters that yield total cost equal to the threshold value,

$$E(x) = \big\{\xi \in \mathbb{R}^N \mid cx + \Phi\big(h(\xi) - T(\xi)x\big) = \phi\big\}.$$

For the subsequent analysis we will also need the following definition of the limes inferior and the limes superior of sequences of sets.

Definition 4.2.1. The limes inferior and the limes superior of a sequence of sets $\{A_n\}_{n=1}^{\infty}$ in $\mathcal{B}(\mathbb{R}^N)$ are

$$\liminf_{n\to\infty} A_n = \bigcup_{j\geq 1} \bigcap_{n\geq j} A_n \qquad \text{and} \qquad \limsup_{n\to\infty} A_n = \bigcap_{j\geq 1} \bigcup_{n\geq j} A_n,$$

respectively.

Hence, given a sequence of sets $\{A_n\}_{n=1}^{\infty}$ in $\mathcal{B}(\mathbb{R}^N)$, the limes inferior is the set of all those $\xi \in \mathbb{R}^N$ for which there exists $n_0 \in \mathbb{N}$ such that $\xi \in A_n$ for all $n \geq n_0$, whereas the limes superior is the set of all those $\xi \in \mathbb{R}^N$ such that $\xi \in A_n$ for infinitely many n. The following continuity property of probability measures is a special case of the so-called Fatou Lemma and may be found for example in Chapter 1.4 in Hoffman-Jørgensen [61].

Lemma 4.2.1. *Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ and let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets in $\mathcal{B}(\mathbb{R}^N)$. Then,*

(a) $\liminf_{n\to\infty} \mu(A_n) \geq \mu\big(\liminf_{n\to\infty} A_n\big)$;

(b) $\limsup_{n\to\infty} \mu(A_n) \leq \mu\big(\limsup_{n\to\infty} A_n\big)$.

We will also need the following lemma.

Lemma 4.2.2. *Assume (C1)-(C2), let $x \in \mathbb{R}^{n_1}$, and let $\{x_n\}_{n=1}^{\infty}$ be some sequence in $\mathbb{R}^{n_1}$ converging to x. Then,*

(a) $\liminf_{n\to\infty} M(x_n) \supseteq M(x)$;

(b) $\limsup_{n\to\infty} M(x_n) \subseteq M(x) \cup E(x)$.

Proof. (a) If $\xi \in M(x)$ we have by definition of $M(x)$ that $cx + \Phi\big(h(\xi) - T(\xi)x\big) > \phi$. By continuity of Φ this means that there exists some $n_0 \in \mathbb{N}$ such that for all $n > n_0$ we have $cx_n + \Phi\big(h(\xi) - T(\xi)x_n\big) > \phi$, and hence we see that $\xi \in M(x_n)$ for all $n > n_0$. The result follows immediately, cf. Definition 4.2.1.
(b) Let $\xi \in \limsup_{n\to\infty} M(x_n) \setminus M(x)$. This means that $cx + \Phi\big(h(\xi) - T(\xi)x\big) \leq \phi$ while $cx_n + \Phi\big(h(\xi) - T(\xi)x_n\big) > \phi$ for infinitely many n, cf. Definition 4.2.1. Now, by continuity of Φ, we see that $cx + \Phi\big(h(\xi) - T(\xi)x\big) = \phi$ and hence $\xi \in E(x)$. □

Lemma 4.2.1 and Lemma 4.2.2 are sufficient to establish the qualitative continuity properties of $\mathcal{Q}_P$ as a function of x, expressed in the following propositions.

Proposition 4.2.1. *Assume (C1)-(C2) and let $\mu \in \mathcal{P}(\mathbb{R}^N)$. Then $\mathcal{Q}_P(\cdot, \mu)$ is a real-valued lower semicontinuous function on $\mathbb{R}^{n_1}$.*

Proof. By continuity of Φ it is easily seen that $M(x)$ is an open set and hence measurable for any $x \in \mathbb{R}^{n_1}$. Thus $\mathcal{Q}_P(\cdot, \mu)$ is well-defined and obviously real-valued. Now, let $x \in \mathbb{R}^{n_1}$ and let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}^{n_1}$ converging to x. By Lemma 4.2.1 (a) and Lemma 4.2.2 (a) we now have

$$\mathcal{Q}_P(x, \mu) = \mu\big(M(x)\big) \leq \mu\big(\liminf_{n\to\infty} M(x_n)\big) \leq \liminf_{n\to\infty} \mu\big(M(x_n)\big) = \liminf_{n\to\infty} \mathcal{Q}_P(x_n, \mu).$$

Hence, $\mathcal{Q}_P(\cdot, \mu)$ is lower semicontinuous at x. □

Proposition 4.2.2. *Assume (C1)-(C2) and let $\mu \in \mathcal{P}(\mathbb{R}^N)$ and $x \in \mathbb{R}^{n_1}$ be such that $\mu\big(E(x)\big) = 0$. Then $\mathcal{Q}_P(\cdot, \mu)$ is continuous at x.*

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}^{n_1}$ converging to x. By the assumption that $\mu\big(E(x)\big) = 0$ we have $\mu\big(M(x)\big) = \mu\big(M(x) \cup E(x)\big)$, and hence by Lemma 4.2.1 (b) and Lemma 4.2.2 (b) we get

$$\mathcal{Q}_P(x, \mu) = \mu\big(M(x)\big) \geq \mu\big(\limsup_{n\to\infty} M(x_n)\big) \geq \limsup_{n\to\infty} \mu\big(M(x_n)\big) = \limsup_{n\to\infty} \mathcal{Q}_P(x_n, \mu).$$

Hence, observing Proposition 4.2.1, we see that $\mathcal{Q}_P(\cdot, \mu)$ is continuous at x. □

Recalling the properties of the marginal and conditional distributions of the second-stage right-hand side and the technology matrix listed in Appendix A, we obtain the following corollary as an immediate consequence of Proposition 4.2.2.

Corollary 4.2.1. *Assume (C1) - (C2) and let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be such that $\mu_1^2(\cdot, T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$ for μ_2-almost all $T \in \mathbb{R}^{m_2 \times n_1}$. Then $\mathcal{Q}_P(\cdot, \mu)$ is a continuous function on $\mathbb{R}^{n_1}$.*

Proof. Let $x \in \mathbb{R}^{n_1}$. (C1) and (C2) imply that by linear programming duality we have

$$\Phi(\tau) = \max_{j \in \{1,\dots,K\}} \big\{ d^j \tau \big\},$$

where $d^1, \dots, d^K$ are the extreme points of the set $\big\{ u \in \mathbb{R}^{m_2} \mid uW \le q \big\}$. Hence the set of all those $\tau \in \mathbb{R}^{m_2}$ such that $cx + \Phi(\tau) = \phi$, is contained in a finite union of hyperplanes $\mathcal{H} = \cup_{j=1}^K H_j$ where $H_j = \big\{ \tau \in \mathbb{R}^{m_2} \mid d^j \tau = \phi - cx \big\}$ for $j = 1, \dots, K$. Thus we see that for any $x \in \mathbb{R}^{n_1}$ we have $E(x) \subseteq \big\{ \xi \in \mathbb{R}^N \mid h(\xi) - T(\xi)x \in \mathcal{H} \big\}$ and hence

$$\begin{aligned} \mu\big(E(x)\big) &= \int_{\mathbb{R}^N} \mathbf{1}_{E(x)}(\xi) \mu(d\xi) = \int_{\mathbb{R}^{m_2 \times n_1}} \int_{\mathbb{R}^{m_2}} \mathbf{1}_{E(x)}(h, T_1, \dots, T_{m_2})\, \mu_1^2(dh, T)\, \mu_2(dT) \\ &\le \int_{\mathbb{R}^{m_2 \times n_1}} \int_{Tx + \mathcal{H}} \mu_1^2(dh, T)\, \mu_2(dT) = 0, \end{aligned}$$

where the last equality follows since the inner integral is equal to zero μ_2-almost surely under the assumption that $\mu_1^2(\cdot, T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$ for μ_2-almost all $T \in \mathbb{R}^{m_2 \times n_1}$. Thus we may apply Proposition 4.2.2 to obtain the desired result. □

Remark 4.2.2. In Chapter 3 we saw that for a classical two-stage stochastic program with mixed-integer recourse, the assumptions of complete (mixed-integer) recourse and dual feasibility of the second-stage problems must be complemented with the additional assumption that the probability distribution μ has finite first moment, in order to arrive at the continuity properties of the expected recourse function expressed in Theorem 3.2.1, Theorem 3.2.2, and Corollary 3.2.1 on page 31. This is in fact a crucial assumption when studying the classical two-stage stochastic program with mixed-integer recourse, required to establish the existence of integrable minorants and majorants of the second-stage value function. From Proposition 4.2.1, Proposition 4.2.2, and Corollary 4.2.1 we see, on the other hand, that in order to obtain the corresponding continuity properties of the recourse function $\mathcal{Q}_P$, it is no longer necessary to assume that μ has finite first moment.

We now turn to the joint continuity of $\mathcal{Q}_P$ as a function of the first-stage decision and the underlying probability measure. To this end, as in the preceding chapters, we will adopt the notion of weak convergence on the set of all Borel probability measures on $\mathbb{R}^N$ (see Appendix A). Extending Remark 4.2.2 we note that in order to establish joint continuity of the expected recourse function of a classical two-stage stochastic program with mixed-integer recourse, using this notion of convergence, one must impose further restrictions on the probability measures, requiring for example uniform boundedness of the integrals $\int_{\mathbb{R}^N} ||\xi||^p \mu(d\xi)$ for some $p > 1$ cf. Theorem 3.2.4 on page 32. According to the following proposition, such a restriction is not necessary for the minimum risk problem.

Proposition 4.2.3. *Assume (C1)-(C2) and let* $\mu \in \mathcal{P}(\mathbb{R}^N)$ *and* $x \in \mathbb{R}^{n_1}$ *be such that* $\mu\big(E(x)\big) = 0$. *Then* $\mathcal{Q}_P$*, as a function from* $\mathbb{R}^{n_1} \times \mathcal{P}(\mathbb{R}^N)$ *to* $\mathbb{R}$*, is continuous at* (x, μ).

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}^{n_1}$ converging to x and let $\{\mu_n\}_{n=1}^{\infty}$ be a sequence in $\mathcal{P}(\mathbb{R}^N)$ converging weakly to μ. Recalling the definition of the indicator function (4.2.2), we now introduce functions $f_n : \mathbb{R}^N \mapsto \mathbb{R}$ and $f : \mathbb{R}^N \mapsto \mathbb{R}$ defined by

$$f_n(\xi) = \psi(x_n, \xi) \quad \text{and} \quad f(\xi) = \psi(x, \xi).$$

Note that all these functions are measurable due to measurability of the open sets $M(x)$ and $M(x_n)$ for $n \geq 1$ cf. Appendix A. Also, we define the set $E_0(x)$ consisting of all those $\xi \in \mathbb{R}^N$ for which there exists a sequence $\{\xi_n\}_{n=1}^{\infty}$ in $\mathbb{R}^N$ such that $\xi_n \to \xi$ but $f_n(\xi_n) \not\to f(\xi)$ as $n \to \infty$.

We now show that $E_0(x) \subseteq E(x)$. So let $\xi \in E_0(x)$ and let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}^N$ converging to ξ such that $f_n(\xi_n) \not\to f(\xi)$ as $n \to \infty$. Assuming that $\xi \in M(x)$, i.e. $cx + \Phi\big(h(\xi) - T(\xi)x\big) > \phi$, we get by continuity of Φ that there exists a number $n_0 \in \mathbb{N}$ such that $\xi_n \in M(x_n)$ for all $n > n_0$ and hence $\lim_{n\to\infty} f_n(\xi_n) = 1 = f(\xi)$, a contradiction. Likewise, the assumption $cx + \Phi\big(h(\xi) - T(\xi)x\big) < \phi$ leads to a contradiction, and hence we must have $\xi \in E(x)$.

Thus, by the assumption $\mu\big(E(x)\big) = 0$ we get $\mu\big(E_0(x)\big) = 0$ and we may apply Rubin's Theorem (see e.g. Billingsley [18, Theorem 5.5]) to obtain

$$\mu_n \circ f_n^{-1} \xrightarrow{w} \mu \circ f^{-1}.$$

Now, because $\mu \circ f^{-1}$ and $\mu_n \circ f_n^{-1}$ are probability measures on $\{0, 1\}$ for all $n \geq 1$, the weak convergence of $\mu_n \circ f_n^{-1}$ to $\mu \circ f^{-1}$ implies in particular that

$$\mu_n \circ f_n^{-1}(\{1\}) \xrightarrow{n\to\infty} \mu \circ f^{-1}(\{1\}),$$

from which we get

$$\mathcal{Q}_P(x_n, \mu_n) = \mu_n\big(M(x_n)\big) \xrightarrow{n\to\infty} \mu\big(M(x)\big) = \mathcal{Q}_P(x, \mu).$$

and the proof is complete. □

Once again, as an immediate consequence of Proposition 4.2.3 and the fact that for all $x \in \mathbb{R}^{n_1}$ the set $E(x)$ is contained in a finite union of hyperplanes, we obtain the following corollary.

Corollary 4.2.2. *Assume (C1) - (C2) and let* $\mu \in \mathcal{P}(\mathbb{R}^N)$ *be such that* $\mu_1^2(\cdot, T)$ *is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^{m_2}$ *for* μ_2*-almost all* $T \in \mathbb{R}^{m_2 \times n_1}$. *Then* $\mathcal{Q}_P$*, as a function from* $\mathbb{R}^{n_1} \times \mathcal{P}(\mathbb{R}^N)$ *to* $\mathbb{R}$*, is continuous on* $\mathbb{R}^{n_1} \times \{\mu\}$.

Proof. The proof is similar to that of Corollary 4.2.1. □

Quantitative continuity of $\mathcal{Q}_P$ as a function of the underlying probability measure μ relies on identifying a (pseudo-) distance on $\mathcal{P}(\mathbb{R}^N)$ that is properly adjusted to the definition of $\mathcal{Q}_P$ via probabilities of level sets of value functions. To this end, denoting for $k \in \mathbb{N}$ by $\mathcal{B}_k \subseteq \mathcal{B}(\mathbb{R}^N)$ the family of all polyhedra in $\mathbb{R}^N$ with at most k faces, the following discrepancy defined for $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$ will turn out useful for our purposes,

$$\alpha_{\mathcal{B}_k}(\mu, \nu) = \sup\big\{|\mu(B) - \nu(B)| \;\big|\; B \in \mathcal{B}_k\big\}. \tag{4.2.6}$$

Remark 4.2.3. As mentioned in Chapter 3 we note that adapting probability (pseudo-) distances to the underlying structures is a proven tool in quantitative stability analysis of stochastic programs. Again we refer to Rachev and Römisch [109] for a general framework and applications to classical recourse models as well as chance-constrained problems.

Proposition 4.2.4. *Assume (C1)-(C2). Then there exists a $k \in \mathbb{N}$ such that for all $x \in \mathbb{R}^{n_1}$ and all $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$ we have*

$$\big|\mathcal{Q}_P(x,\mu) - \mathcal{Q}_P(x,\nu)\big| \leq \alpha_{\mathcal{B}_k}(\mu,\nu).$$

Proof. Denoting by $M^c(x)$ the complement of $M(x)$ we have for any $x \in \mathbb{R}^{n_1}$

$$\begin{aligned} M^c(x) &= \big\{\xi \in \mathbb{R}^N \mid cx + \Phi\big(h(\xi) - T(\xi)x\big) \leq \phi\big\} \\ &= \big\{\xi \in \mathbb{R}^N \mid d^j h(\xi) - d^j T(\xi)x \leq \phi - cx,\ j = 1, \ldots, K\big\}, \end{aligned}$$

where, once again, $d^1, \ldots, d^K$ are the extreme points of the set $\big\{u \in \mathbb{R}^{m_2} \mid uW \leq q\big\}$. Recalling that for $\xi \in \mathbb{R}^N$ we have $\xi = \big(h(\xi), T_1(\xi), \ldots, T_{m_2}(\xi)\big)$, where T_i denotes the ith row of T for $i = 1, \ldots, m_2$, it follows that $\{M^c(x)\}_{x\in\mathbb{R}^{n_1}}$ is a family of polyhedra in $\mathbb{R}^N$ whose numbers of facets are bounded above by a uniform constant, i.e. a constant not depending on x. Hence, there exists a $k \in \mathbb{N}$ such that

$$\begin{aligned} \big|\mathcal{Q}_P(x,\mu) - \mathcal{Q}_P(x,\nu)\big| &= \big|\mu(M(x)) - \nu(M(x))\big| \\ &= \big|\mu(M^c(x)) - \nu(M^c(x))\big| \\ &\leq \sup\big\{|\mu(B) - \nu(B)| \,\big|\, B \in \mathcal{B}_k\big\}, \end{aligned}$$

and the proof is complete. □

Remark 4.2.4. The discrepancy $\alpha_{\mathcal{B}_k}$ is easily seen to be a pseudometric on $\mathcal{P}(\mathbb{R}^N)$, majorizing the discrepancy $\alpha_{\mathcal{B}_\mathcal{K}}$ defined by (3.2.1) on page 32 in connection with the classical stochastic program with mixed-integer recourse. Using the alternative discrepancy $\alpha_{\mathcal{B}_k}$ here, we see that the result of Theorem 3.2.5 on page 33 is extended for the minimum risk problem by Proposition 4.2.4, holding also for random technology matrix. Still, as with the discrepancy $\alpha_{\mathcal{B}_\mathcal{K}}$, a coherence between the discrepancy $\alpha_{\mathcal{B}_k}$ and weak convergence of probability measures may be established using the concept of a μ-uniformity class. Hence we recall that for some probability measure $\mu \in \mathcal{P}(\mathbb{R}^N)$, a family of Borel sets $\mathcal{B}_0 \subseteq \mathcal{B}(\mathbb{R}^N)$ is called a μ-uniformity class if $\sup\big\{|\mu_n(B) - \mu(B)| \,\big|\, B \in \mathcal{B}_0\big\} \xrightarrow{n\to\infty} 0$ for every sequence of probability measures $\{\mu_n\}_{n=1}^\infty$ in $\mathcal{P}(\mathbb{R}^N)$ converging weakly to μ. Denoting by $\mathcal{B}_c$ the class of all convex Borel sets in $\mathbb{R}^N$, we obviously have $\mathcal{B}_k \subseteq \mathcal{B}_c$, and since it was shown by Bhattacharya and Ranga Rao [15, Theorem 2.11], that $\mathcal{B}_c$ is in fact a μ-uniformity class for all those $\mu \in \mathcal{P}(\mathbb{R}^N)$ that are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^N$, we see that $\alpha_{\mathcal{B}_k}(\mu_n,\mu) \xrightarrow{n\to\infty} 0$ for any sequence of probability measures $\{\mu_n\}_{n=1}^\infty$ in $\mathcal{P}(\mathbb{R}^N)$ converging weakly to such μ. Thus, if μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^N$ and $\mu_1^2(\cdot,T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$ for μ_2-almost all $T \in \mathbb{R}^{m_2\times n_1}$, then Proposition 4.2.4 may be seen as a quantification of the result in Corollary 4.2.2. Furthermore, as we will discuss in Section 4.2.3, the class $\mathcal{B}_k$ is in fact a Vapnik-Červonenkis class. This fact will allow us for example to derive conclusions on the asymptotic convergence of optimal solutions for the minimum risk problem when the probability measure μ is estimated by empirical measures.

4.2.3 Stability

As in the preceding chapters let us note that in many practical applications of stochastic programming, the probability distribution of random parameters will not be completely known, and hence the true distribution may have to be replaced in the model by some suitable estimate, such as e.g. empirical measures. Furthermore, even if the true distribution μ of random parameters is known, the approximation of μ by simpler probability measures may be required to facilitate practical computations. In fact, the solution procedure for the minimum risk problem that we will elaborate in Section 4.3, relies on the assumption that the probability distribution of random parameters is discrete with finite support, which will obviously often not be the case in practice. Therefore, just as for the classical stochastic recourse problems considered in Chapters 2 and 3, stability of the minimum risk problem (4.1.1), when the underlying probability distribution is subjected to perturbations, is an important issue to which we now turn.

Since the minimum risk problem $MRP(\mu)$, defined by (4.1.1)-(4.1.3), is in general a non-convex problem, local minimizers are included in the analysis. To this end, we introduce for any non-empty open set $V \subseteq \mathbb{R}^{n_1}$ a localized version, $\varphi_V : \mathcal{P}(\mathbb{R}^N) \mapsto \mathbb{R}$, of the optimal-value function, defined by

$$\varphi_V(\mu) = \inf\big\{\mathcal{Q}_P(x,\mu) \mid Ax = b,\ x \in X \cap \operatorname{cl} V\big\},$$

and a localized version, $\Psi_V : \mathcal{P}(\mathbb{R}^N) \mapsto \mathbb{R}^{n_1}$, of the solution set mapping, defined by

$$\Psi_V(\mu) = \arg\min\big\{\mathcal{Q}_P(x,\mu) \mid Ax = b,\ x \in X \cap \operatorname{cl} V\big\},$$

where $\operatorname{cl} V$ denotes the closure of V.

Remark 4.2.5. Note that if $\big\{x \in X \cap \operatorname{cl} V \mid Ax = b\big\}$ is non-empty and bounded, the infimum in the definition of φ_V is always attained since we are minimizing a lower semicontinuous function over a compact set, and hence in this case $\Psi_V(\mu)$ is non-empty for any $\mu \in \mathcal{P}(\mathbb{R}^N)$.

Having established the joint continuity of $\mathcal{Q}_P$ with respect to x and μ, it is again straightforward to follow the lines of Berge [14] to prove continuity of φ and Berge upper semicontinuity of Ψ, cf. also the results of Bank et al. [9]. (Recall that the point-to-set mapping, Ψ_V, is Berge upper semicontinuous at some $\mu \in \mathcal{P}(\mathbb{R}^N)$ if for any open set $G \subseteq \mathbb{R}^{n_1}$ with $\Psi_V(\mu) \subseteq G$ there exists some neighborhood U of μ in $\mathcal{P}(\mathbb{R}^N)$ such that $\Psi_V(\nu) \subseteq G$ for all $\nu \in U$.) As illustrated by Example 3.2.1 on page 34 for a classical stochastic program with mixed-integer recourse, however, it may happen that $\Psi_V(\mu)$ is a set of local minimizers of $MRP(\mu)$ for some $\mu \in \mathcal{P}(\mathbb{R}^N)$ and $V \subseteq \mathbb{R}^{n_1}$, while for any neighborhood U of μ in $\mathcal{P}(\mathbb{R}^N)$ there exists $\nu \in U$ such that $\Psi_V(\nu)$ does not contain any local minimizers of $MRP(\nu)$. Therefore we will not find the continuity of φ and the Berge upper semicontinuity of Ψ quite sufficient in their own for the stability analysis of local minimizers.

As discussed in Section 3.2.2, Robinson [121] and Klatte [73] proposed a local stability analysis for non-convex problems, precluding pathologies as those mentioned above. Emphasizing the need for considerations to include all local minimizers that are, in some sense, nearby the minimizers one is interested in, the authors introduced the crucial concept of a complete local minimizing (CLM) set, defined as follows. Let μ be a Borel probability measure and let M be a non-empty subset of $\mathbb{R}^{n_1}$. If there exists an open set $V \subseteq \mathbb{R}^{n_1}$ such that $M \subseteq V$ and $M = \Psi_V(\mu)$, then M is called a CLM set for $MRP(\mu)$ with respect to V. Obvious examples of CLM sets are the set of global minimizers as well as any set of strict local minimizers. Hence, the subsequent propositions stated in general for CLM sets are valid in particular for the set of global minimizers and for any set of strict local minimizers.

Proposition 4.2.5. *Assume (C1)-(C2), let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be such that $\mu_1^2(\cdot, T)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$ for μ_2-almost all $T \in \mathbb{R}^{m_2 \times n_1}$, and let $V \subseteq \mathbb{R}^{n_1}$ be a bounded open set such that $\Psi_V(\mu)$ is a CLM set for $MRP(\mu)$ with respect to V. Then,*

(a) $\varphi_V : \mathcal{P}(\mathbb{R}^N) \mapsto \mathbb{R}$ is continuous at μ;

(b) $\Psi_V : \mathcal{P}(\mathbb{R}^N) \mapsto \mathbb{R}^{n_1}$ is Berge upper semicontinuous at μ;

(c) there exists some neighborhood U of μ in $\mathcal{P}(\mathbb{R}^N)$ such that $\Psi_V(\nu)$ is a CLM set for $MRP(\nu)$ with respect to V for all $\nu \in U$.

Proof. (a) Continuity of φ_V is an immediate consequence of Corollary 4.2.2 and compactness of $X \cap \operatorname{cl} V$, cf. e.g. the proof of Theorem 4.2.2 in Bank et al. [9].
(b) Let $\{\mu_n\}_{n=1}^{\infty}$ be a sequence in $\mathcal{P}(\mathbb{R}^N)$ converging weakly to μ and let $x_n \in \Psi_V(\mu_n)$ for all $n \geq 1$ such that the sequence $\{x_n\}_{n=1}^{\infty}$ converges to some $x \in \mathbb{R}^{n_1}$. By Corollary 4.2.2 and continuity of φ_V we now have

$$\mathcal{Q}_P(x, \mu) = \lim_{n\to\infty} \mathcal{Q}_P(x_n, \mu_n) = \lim_{n\to\infty} \varphi_V(\mu_n) = \varphi_V(\mu).$$

Thus $x \in \Psi_V(\mu)$ implying that Ψ_V is a closed mapping and hence Berge upper semicontinuous by compactness of $X \cap \operatorname{cl} V$, cf. Lemma 2.2.3 in Bank et al. [9].
(c) By Berge upper semicontinuity of Ψ_V there exists some neighborhood U of μ in $\mathcal{P}(\mathbb{R}^N)$ such that $\Psi_V(\nu) \subseteq V$ for all $\nu \in U$. Non-emptiness of $\Psi_V(\nu)$ for $\nu \in U$ follows from non-emptiness of $\Psi_V(\mu)$, boundedness of V, and lower semicontinuity of $\mathcal{Q}_P(\cdot, \nu)$. □

Once again we may quantify the result in Proposition 4.2.5 (a) using the pseudometric $\alpha_{\mathcal{B}_k}$ that was defined by (4.2.6) on page 52.

Proposition 4.2.6. *Assume (C1)-(C2), let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^N$, and let $V \subseteq \mathbb{R}^{n_1}$ be a bounded open set such that $\Psi_V(\mu)$ is a CLM set for $MRP(\mu)$ with respect to V. Then there exists a $k \in \mathbb{N}$ such that*

$$\left|\varphi_V(\mu) - \varphi_V(\nu)\right| \leq \alpha_{\mathcal{B}_k}(\mu, \nu).$$

for all $\nu \in \mathcal{P}(\mathbb{R}^N)$.

Proof. Let $\nu \in \mathcal{P}(\mathbb{R}^N)$ and note once again that $\Psi_V(\nu)$ is non-empty by non-emptiness of $\Psi_V(\mu)$, boundedness of V, and lower semicontinuity of $\mathcal{Q}_P(\cdot, \nu)$. Now, let $x_\mu \in \Psi_V(\mu)$

and $x_\nu \in \Psi_V(\nu)$, and apply Proposition 4.2.4 to obtain the existence of some $k \in \mathbb{N}$ such that

$$\varphi_V(\mu) - \varphi_V(\nu) \le \left|\mathcal{Q}_P(x_\nu, \mu) - \mathcal{Q}_P(x_\nu, \nu)\right| \le \alpha_{\mathcal{B}_k}(\mu, \nu)$$

and

$$\varphi_V(\nu) - \varphi_V(\mu) \le \left|\mathcal{Q}_P(x_\mu, \nu) - \mathcal{Q}_P(x_\mu, \mu)\right| \le \alpha_{\mathcal{B}_k}(\mu, \nu).$$

Hence,

$$\left|\varphi_V(\mu) - \varphi_V(\nu)\right| \le \alpha_{\mathcal{B}_k}(\mu, \nu),$$

and the proof is complete. □

Remark 4.2.6. Here the assumption that μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^N$, is required only to ensure that $\mathcal{B}_k$ is a μ-uniformity class cf. Remark 4.2.4, hence implying that $\alpha_{\mathcal{B}_k}(\mu_n, \mu) \xrightarrow{n\to\infty} 0$ for any sequence of probability measures $\{\mu_n\}_{n=1}^\infty$ in $\mathcal{P}(\mathbb{R}^N)$ converging weakly to μ.

As previously pointed out, the true distribution of random parameters will not be completely known in many practical applications of stochastic programming, and hence the true distribution may have to be replaced by some suitable estimate. Let us consider for a moment the situation when the true distribution μ is approximated by empirical measures. In particular, we let $\{\xi_n\}_{n=1}^\infty$ be a sequence of independent and identically distributed N-dimensional random vectors defined on some probability space $(\Omega, \mathcal{F}, P)$, and we denote by μ their common distribution. This gives rise to a corresponding sequence of empirical probability measures on $\mathbb{R}^N$ defined by

$$\mu_n(\omega) = \frac{1}{n}\sum_{i=1}^n \delta_{\xi_i(\omega)},$$

where $\delta_{\xi_i(\omega)}$ denotes the measure with unit mass at $\xi_i(\omega)$ for $i = 1, \ldots, n$. It is well-known that we have $\mu_n(\omega) \xrightarrow{w} \mu$ for P-almost all $\omega \in \Omega$ cf. e.g. Dudley [42, Theorem 11.4.1]. As in the linear recourse case one may now apply Proposition 4.2.5 to obtain asymptotic convergence of local optimal values and local optimal solutions when μ satisfies the hypotheses of that theorem. Following the lines of Schultz [141, page 1148], though, it is easily seen that the class $\mathcal{B}_k$ is a Vapnik-Červonenkis class, i.e. there exists some number $r \in \mathbb{N}$ such that for any finite set $E \subseteq \mathbb{R}^N$ with r elements, not all subsets of E are of the form $E \cap B$, $B \in \mathcal{B}_k$. Furthermore, it is straightforward to extend the arguments of Schultz to show measurability of $\alpha_{\mathcal{B}_k}(\mu_n(\omega), \mu)$ as a function of ω, and hence well-known results may be applied to show the following.

Lemma 4.2.3. *Let $k \in \mathbb{N}$. Then $\alpha_{\mathcal{B}_k}(\mu_n(\omega), \mu) \xrightarrow{n\to\infty} 0$ for P-almost all $\omega \in \Omega$.*

Proof. For details we refer to Schultz [141]. □

As in Schultz [141] we may now apply Proposition 4.2.4 to obtain the following result, where, in particular, the smoothness assumptions on μ required in Proposition 4.2.5 and Proposition 4.2.6 are abandoned.

Proposition 4.2.7. *Assume (C1)-(C2), let* $\mu \in \mathcal{P}(\mathbb{R}^N)$*, and let* $V \subseteq \mathbb{R}^{n_1}$ *be a bounded open set such that* $\Psi_V(\mu)$ *is a CLM set for* $MRP(\mu)$ *with respect to* V*. Then,*

(a) $\varphi_V(\mu_n(\omega)) \xrightarrow{n\to\infty} \varphi_V(\mu)$ *for* P*-almost all* $\omega \in \Omega$*;*

(b) for any open set $G \subseteq \mathbb{R}^{n_1}$ *with* $\Psi_V(\mu) \subseteq G$ *and for* P*-almost all* $\omega \in \Omega$ *there exists some* $n_0(\omega) \in \mathbb{N}$ *such that* $\Psi_V(\mu_n(\omega)) \subseteq G$ *for all* $n \geq n_0(\omega)$*;*

(c) for P*-almost all* $\omega \in \Omega$ *there exists some* $n_1(\omega) \in \mathbb{N}$ *such that* $\Psi_V(\mu_n(\omega))$ *is a CLM set for* $MRP(\mu_n(\omega))$ *with respect to* V *for all* $n \geq n_1(\omega)$*.*

Proof. The proof closely follows that of Schultz [141, Proposition 5.3]. □

Remark 4.2.7. Comparing the stability results of Propositions 4.2.5, 4.2.6, and 4.2.7 with those presented for the general classical stochastic program with mixed-integer recourse in Theorems 3.2.6, 3.2.7, and 3.2.8 on pages 35-36, we see once again that the special structure of the minimum risk problem allows us to abandon the integrability assumptions imposed in the analysis of the classical recourse problem. Also, the quantitative stability results are extended for the minimum risk problem in the sense that they now hold for problems with random technology matrix too, cf. also Remark 4.2.4.

4.3 Solution Procedure

In this section we elaborate a specialized solution procedure for the minimum risk problem. For practical purposes we need to make the following assumptions.

(C3) The first-stage solution set $\{x \in X \mid Ax = b\}$ is non-empty and compact.

(C4) The distribution μ of ξ is discrete and has finite support, say $\Xi = \{\xi^1, \ldots, \xi^S\}$ with corresponding probabilities $p^1, \ldots, p^S$.

For each $s \in \{1, \ldots, S\}$, a possible outcome $\big(h(\xi^s), T(\xi^s)\big)$ of random parameters, corresponding to some elementary event $\xi^s \in \Xi$, is referred to as a *scenario*, and we denote it simply by (h^s, T^s).

Remark 4.3.1. Note that (C4) may be justified by the stability results established in the previous section. Thus, according to Proposition 4.2.5 for example, the optimal value and the solution set of a problem with a continuous distribution of random parameters may be approximated to any given accuracy by the optimal value and the solution set of problems employing only discrete distributions.

As pointed out in Section 4.2.1, the minimum risk problem is under assumptions (C3) and (C4) equivalent to a classical stochastic program with mixed-integer recourse, where the expected value of an appropriately defined indicator function is minimized. A possible way to go is therefore to define such an indicator function by (4.2.4), including a binary variable and an additional constraint in the second stage, and subsequently to solve the problem by one of the solution procedures developed for the general class of two-stage stochastic programs with mixed-integer recourse, cf. e.g. the discussion in Section 3.3. As we have seen in the preceding chapters, however, the problem complexity drastically increases when a binary variable is included in the second stage. In particular, the original second-stage value function (4.1.3) is piecewise linear and convex, whereas the transformed one (4.2.4) would lose these appealing properties. In this section we show how one may avoid the inclusion of a binary second-stage variable by solving the minimum risk problem using a modified version of the L-shaped procedure (see Section 2.3.1). The idea is for each scenario $s \in \{1, \dots, S\}$ to represent the general indicator function $\psi(\cdot, \xi^s)$, defined by (4.2.2), by a binary variable and a number of optimality cuts, similar in vein to those used for ordinary two-stage stochastic linear programs. By not applying the definition of the indicator function given by (4.2.4), we obtain a formulation in which binary variables occur only in a master problem and not in the second-stage subproblems.

Given a first-stage solution $x \in X$ and a scenario $s \in \{1, \dots, S\}$, the optimality cuts needed to represent $\psi(\cdot, \xi^s)$ at x are derived from the linear programming problem,

$$\begin{aligned} \min\ & et^+ + et^- + t_0 \\ \text{s.t. } & Wy + It^+ - It^- = h^s - T^s x, \\ & qy - t_0 \le \phi - cx, \\ & y \in \mathbb{R}_+^{n_2},\ t^+, t^- \in \mathbb{R}_+^{m_2},\ t_0 \in \mathbb{R}_+, \end{aligned} \tag{4.3.1}$$

and its dual problem

$$\begin{aligned} \max\ & (h^s - T^s x)u + (cx - \phi)u_0 \\ \text{s.t. } & uW - u_0 q \le 0, \\ & -e \le Iu \le e,\ u_0 \le 1, \\ & u \in \mathbb{R}^{m_2},\ u_0 \in \mathbb{R}_+, \end{aligned} \tag{4.3.2}$$

where $e = (1, \dots, 1) \in \mathbb{R}^{m_2}$ and I is the $m_2 \times m_2$-identity matrix. Note that for all $x \in \mathbb{R}^{n_1}$ and all $s \in \{1, \dots, S\}$, both the primal problem (4.3.1) and the dual problem (4.3.2) are feasible, and hence they are both solvable. Moreover, their optimal values are identical and non-negative, and equal to zero if and only if $\psi(x, \xi^s) = 0$. In the following we will denote by D the feasible region of the dual problem (4.3.2), and by $(d^1, d_0^1), \dots, (d^K, d_0^K)$ the extreme points of D. Also, we let $M_1 > 0$ be some large number bounding from above the optimal value of the dual problem,

$$M_1 = \sup\Big\{ (h^s - T^s x)u + (cx - \phi)u_0 \,|\, Ax = b,\ x \in X,\ (u, u_0) \in D,\ s \in \{1, \dots, S\} \Big\}.$$

Note once again that, under assumptions (C3) and (C4), the supremum exists and is finite, since D is obviously bounded.

The following lemmas, elucidating the structure of the optimality cuts, are immediate consequences of the definition of M_1 and the previously discussed relationship between the indicator function (4.2.2) and the linear programming problems (4.3.1) and (4.3.2).

Lemma 4.3.1. *For all $x \in X$ such that $Ax = b$, and for any scenario $s \in \{1, \dots. S\}$, the indicator function $\psi(x, \xi^s)$ satisfies the following set of inequalities,*

$$M_1 \psi(x, \xi^s) \ge (h^s - T^s x)d^j + (cx - \phi)d_0^j, \quad j = 1, \dots, K.$$

Lemma 4.3.2. *Let $x \in X$ be such that $\psi(x, \xi^s) = 1$ for some scenario $s \in \{1, \dots. S\}$. Then there exists $j \in \{1, \dots, K\}$ such that $(h^s - T^s x)d^j + (cx - \phi)d_0^j > 0$.*

Observing Lemma 4.3.1 and Lemma 4.3.2 it is easily seen that the minimum risk problem is equivalent to the following mixed-integer program,

$$\begin{aligned} \min \ & \sum_{s=1}^{S} p^s \theta^s \\ \text{s.t. } & Ax = b, \\ & (h^s - T^s x)d^j + (cx - \phi)d_0^j \leq M_1 \theta^s, \quad j = 1, \dots, K, \ s = 1, \dots, S, \\ & x \in X, \ \theta^1, \dots, \theta^S \in \{0, 1\}. \end{aligned} \tag{4.3.3}$$

The algorithm progresses by sequentially solving a master problem and adding violated optimality cuts generated through the solution of subproblems (4.3.1)-(4.3.2).

Algorithm 4.1

Step 1 (*Initialization*) Set $\nu = 0$, and let the current master problem be defined by $\min\{\sum_{s=1}^{S} p^s \theta^s \mid Ax = b, \ x \in X, \ \theta^1, \dots, \theta^S \in \{0, 1\}\}$.

Step 2 (*Solve master problem*) Set $\nu = \nu + 1$. Solve the current master problem and let $(x^\nu, \theta^{1,\nu}, \dots, \theta^{S,\nu})$ be an optimal solution vector.

Step 3 (*Solve subproblems*) Solve the second-stage problem (4.3.1)-(4.3.2) for all $s \in \{1, \dots, S\}$ such that $\theta^{s,\nu} = 0$. Consider the following situations,

(i) if all of these problems have optimal value equal to zero, stop — the current solution x^ν is optimal;

(ii) if some of these problems have optimal value greater than zero, then an equal number of dual extreme points (d^j, d_0^j), $j \in \mathcal{K} \subseteq \{1, \dots, K\}$, each of which satisfies $(h^s - T^s x^\nu)d^j + (cx^\nu - \phi)d_0^j > 0$, are identified and the corresponding optimality cuts are added to the master; go to Step 2.

It is easily seen that Algorithm 4.1 terminates finitely.

Proposition 4.3.1. *Assume (C1)-(C4). Then Algorithm 4.1 terminates with an optimal solution in a finite number of iterations.*

Proof. By assumption (C3) and Proposition 4.2.1 an optimal solution of the minimum risk problem exists. Let x^* be one such solution, and denote by z^* the optimal value. First of all note that the optimal value z^ν of the master problem in iteration ν is a lower bound on z^*, since the master problem is a relaxation of (4.3.3), i.e. for all $\nu \geq 1$ we have

$$\sum_{s=1}^{S} p^s \theta^{s,\nu} = z^\nu \leq z^* = \sum_{s=1}^{S} p^s \psi(x^*, \xi^s).$$

Now, if $\theta^{s,\nu} < \psi(x^\nu, \xi^s)$ for some $s \in \{1, \dots, S\}$ and $\nu \geq 1$, a violated optimality cut, cutting off the current solution $(x^\nu, \theta^{1,\nu}, \dots, \theta^{S,\nu})$, is identified in Step 3 cf. Lemma 4.3.2, and the algorithm proceeds. This can only happen a finite number of times because the number of dual extreme points is finite, and hence we will eventually have

$$\theta^{s,\nu} \geq \psi(x^\nu, \xi^s), \quad s = 1, \dots, S,$$

at which point the current solution x^ν is optimal, and the algorithm terminates. $\square$

Remark 4.3.2. Note that Algorithm 4.1 works equally well for problems which do not satisfy the (relatively) complete recourse property. The optimal first-stage solution determined by the algorithm, however, may not guarantee feasibility for all second-stage problems in this case. Still, the algorithm may easily be modified to accommodate the possible requirement of feasibility of all second-stage problems by using feasibility cuts as in the original L-shaped algorithm.

4.3.1 Solving Mean-Risk Problems

Let us consider again the mean-risk model discussed in Section 4.1.1, where the expected cost and the probability of total cost exceeding the threshold value are simultaneously minimized. We recall that this model allows a simple tradeoff analysis between the two objectives, involving for example the solution of mean-risk problems of the form

$$\min\bigl\{cx + \mathcal{Q}_E(x,\mu) + \alpha\mathcal{Q}_P(x,\mu) \mid Ax = b,\ x \in X\bigr\}, \tag{4.3.4}$$

where $\mathcal{Q}_E(x,\mu) = \mathbb{E}\bigl[\Phi\bigl(h(\xi) - T(\xi)x\bigr)\bigr]$ is the classical expected recourse function considered in Chapter 2, and $\alpha > 0$ is a weight factor.

Observing the similarity between Algorithm 4.1 and the L-shaped decomposition procedure for classical stochastic programs with linear recourse (Algorithm 2.1 on page 25), it seems obvious that a hybrid version of the two algorithms for the solution of the mean-risk problem (4.3.4) is appropriate when the probability distribution μ is discrete with finite support. In such a hybrid solution scheme, the probability-based recourse function $\mathcal{Q}_P$ is approximated by means of optimality cuts derived from the primal-dual pair of problems (4.3.1)-(4.3.2) as described above, and likewise the expected recourse function $\mathcal{Q}_E$ is approximated by means of optimality cuts derived from the second-stage problem (4.1.3) and its dual as discussed in Section 2.3.1.

Obviously, when formalizing such a hybrid solution procedure, some effort should be made to take advantage of the similarity between the second-stage problem (4.1.3) and the subproblems (4.3.1). To this end, it turns out that the solution of (4.3.1)-(4.3.2) is in fact superfluous, because optimality cuts for the indicator function ψ defined by (4.2.2), can be derived from the solution of the second-stage problem and the fact that for $x \in \mathbb{R}^{n_1}$ and $s \in \{1,\dots,S\}$ we have $\psi(x,\xi^s) = 1$ if and only if $\Phi(h^s - T^s x) > \phi - cx$. In particular, given a first-stage solution $x \in X$, one must solve for all scenarios $s \in \{1,\dots,S\}$, the second-stage problem

$$\Phi(h^s - T^s x) = \min\bigl\{q^s y \mid Wy = h^s - T^s x,\ y \in \mathbb{R}^{n_2}_+\bigr\}, \tag{4.3.5}$$

and the corresponding dual problem

$$\Phi(h^s - T^s x) = \max\bigl\{(h^s - T^s x)u \mid uW \le q,\ u \in \mathbb{R}^{m_2}\bigr\}, \tag{4.3.6}$$

to generate optimality cuts representing $\mathcal{Q}_E(\cdot,\mu)$ at x. Now, if $\Phi(h^s - T^s x) \le \phi - cx$ for some scenario $s \in \{1,\dots,S\}$, we see that $\psi(x,\xi^s) = 0$, and hence no optimality cuts are required to represent $\psi(\cdot,\xi^s)$ at x. Therefore, in this case we may immediately conclude that the subproblems (4.3.1)-(4.3.2) does not have to be solved. If, on the other hand, $\Phi(h^s - T^s x) > \phi - cx$ for some scenario $s \in \{1,\dots,S\}$, we see that $\Psi(x,\xi^s) = 1$, in which case a feasible solution (u, u_0) of (4.3.2) with $(h^s - T^s x)u + (cx - \phi)u_0 > 0$ is required to form an optimality cut for $\psi(\cdot, \xi^s)$ at x. As suggested by the following lemma, however, such a dual solution may be generated from a solution of (4.3.6), and hence the subproblems (4.3.1)-(4.3.2) does not have to be solved in this case either.

Lemma 4.3.3. *Let $s \in \{1, \ldots, S\}$, let $u^* \in \mathbb{R}^{m_2}$ be a feasible solution of (4.3.6) with $(h^s - T^s x)u^* > \phi - cx$, and define $U = \max\{1, |u_1^*|, \ldots, |u_{m_2}^*|\}$. Then*

$$(u, u_0) = \frac{1}{U}(u^*, 1)$$

is a feasible solution of (4.3.2) with $(h^s - T^s x)u + (cx - \phi)u_0 > 0$.

Proof. By construction we have $1/U \leq 1$ and $-1 \leq u_k^*/U \leq 1$ for $k = 1, \ldots, m_2$, and since u^* is feasible for (4.3.6) we have $u^*W \leq q$ and hence $uW - u_0 q = \frac{1}{U}(u^*W - q) \leq 0$, showing that (u, u_0) is feasible for (4.3.2). Moreover, since $(h^s - T^s x)u^* > \phi - cx$ and $U > 0$ we have $(h^s - T^s x)u + (cx - \phi)u_0 = \big((h^s - T^s x)u^* + cx - \phi\big)/U > 0$. $\square$

4.4 Computational Experiments

Algorithm 4.1 was implemented in C++ using procedures from the callable library of CPLEX 7.0. Because of its similarity with the L-shaped decomposition procedure, the algorithm is bound to suffer from some of the same drawbacks. Of particular importance in this respect, is the fact that early iterations will usually be quite inefficient since solutions tend to oscillate heavily. This deplorable behavior may be surmounted if a regularizing term, penalizing divergence from the current solution, is added to the master objective, cf. the regularized decomposition procedure introduced by Ruszczyński [132] and discussed in Section 2.3.2. Furthermore, adding a regularizing term to the master objective potentially allows the algorithm to take advantage of a starting solution x^0. Regularized decomposition, as well as most bundle methods for nonsmooth optimization, gain considerable advantage by using a quadratic regularizing term. To avoid a nonlinear mixed-integer formulation of the master problem, however, we simply added to the objective, a term of the form $\gamma e y$, where $\gamma > 0$ is a scaling factor, $e = (1, \ldots, 1) \in \mathbb{R}^{n_1}$, and the variable $y \in \mathbb{R}_+^{n_1}$ represents the deviation from an incumbent solution $a^\nu \in \mathbb{R}^{n_1}$, i.e. we added the constraints

$$y \geq x - a^\nu \quad \text{and} \quad y \geq a^\nu - x.$$

to the master problem in iteration ν. Because of the mixed-integer nature of the minimum risk problem, the inclusion of the regularizing term is not theoretically justified as in regularized decomposition, in the sense that convergence of optimal solutions of the regularized master problem to an optimal solution of the minimum risk problem, cannot be established in general. In practice, we chose to start off the algorithm with the regularizing term included in the objective and remove the regularization once solutions had stabilized. In most cases the algorithm terminated with an optimal solution after just one additional iteration but on a few occasions several iterations had to be performed after removing the regularization.

To investigate the practicability of Algorithm 4.1, we used three sets of problem instances subsequently referred to as EPS1, EPS2, and EPS3, respectively. The problems were obtained as linear programming relaxations of certain mixed-integer programs arising as scheduling problems in chemical production. (See Engell et al. [47] for details.) We ran the algorithm with varying number of scenarios S as well as varying values of the threshold value ϕ. Each run was performed with two different versions of the algorithm. In the following, MRP refers to the algorithm as presented in Section 4.3 while MRPREG refers to the algorithm using the regularization of the master problem described above. At termination of each run we recorded the optimal value, the number of iterations performed, the number of generated cuts and the CPU time spent by the procedure (in seconds). Finally, we fixed the first-stage variables at their values from the optimal solution of the corresponding classical stochastic program to calculate the objective value of that solution, referred to as VCSP in the following. All computational experiments were carried out on a SUN Enterprise 450, 300 MHz Ultra-SPARC.

Let us first consider the EPS1 instance. This problem contains 2 constraints and 3 variables in the first stage and 50 constraints and 51 variables in the second stage. For this instance we ran the algorithm with 20, 50, 100, 200, and 500 scenarios, and each time we chose the threshold value close to the optimal value of the corresponding classical stochastic program. Results are reported in Table 4.1.

Table 4.1: Computational results for the EPS1 instance

				MRP			MRPREG		
S	ϕ	Opt.	VCSP	Ite.	Cuts	CPU	Ite.	Cuts	CPU
20	18.9	0.250	0.250	5	73	0.31	6	58	0.29
50	50.9	0.120	0.140	5	110	0.51	7	110	0.63
100	32.7	0.860	0.860	4	300	0.88	6	300	0.93
200	28.4	0.115	1.000	8	940	8.15	8	761	6.55
500	757.5	0.048	0.048	5	1518	6.56	6	1518	6.62

Next we turn to the EPS2 instance. This problem contains 5 constraints and 12 variables in the first stage and 157 constraints and 164 variables in the second stage. For this instance we always used 100 scenarios and solved the problem for a number of different threshold values surrounding the optimal value of the corresponding classical stochastic program, which was 65.4. Results are reported in Table 4.2.

Table 4.2: Computational results for the EPS2 instance

				MRP			MRPREG		
S	ϕ	Opt.	VCSP	Ite.	Cuts	CPU	Ite.	Cuts	CPU
100	60	0.91	0.92	19	851	36.98	14	728	22.93
100	63	0.81	0.81	19	1098	61.11	15	920	28.51
100	65	0.71	0.71	18	1146	196.96	14	868	53.77
100	67	0.17	0.17	24	1811	34.46	12	880	8.54
100	70	0.08	0.08	18	1322	12.12	13	782	10.50

Finally, we consider the EPS3 instance. This problem contains 9 constraints and 30 variables in the first stage and 280 constraints and 326 variables in the second stage. Once again, we always used 100 scenarios and solved the problem for a number of different threshold values surrounding the optimal value of the corresponding classical stochastic program, which in this case was 191.3. Results are reported in Table 4.3.

Table 4.3: Computational results for the EPS3 instance

				MRP			MRPREG		
S	ϕ	Opt.	VCSP	Ite.	Cuts	CPU	Ite.	Cuts	CPU
100	170	0.94	0.97	41	1407	8164.01	46	1093	1858.37
100	180	0.85	0.88	45	1418	5564.62	30	1326	4632.01
100	190	0.59	0.60	37	1996	4716.74	25	1328	2019.36
100	200	0.13	0.14	44	2317	178.90	24	1280	88.26
100	210	0.02	0.02	62	2656	142.89	23	1455	36.30

We note that the optimal value of the corresponding classical stochastic program of the EPS1 instance with 200 scenarios was 28.7 with all scenarios having cost above 28.4, and hence the seemingly strange result for this instance reported in Table 4.1 was obtained. Apart from this instance, however, the value of the classical stochastic programming solution is always relatively close to the optimal value of the minimum risk problem, and hence the gain of solving the minimum risk problem rather than the corresponding classical stochastic program is negligible for the instances considered here. We did, however, also test the algorithm on the linear programming relaxation of a small stochastic program, previously used as test instance in papers by Carøe and Schultz [29] and Schultz, Stougie, and van der Vlerk [143], and for this problem the gain of solving the minimum risk problem was more significant as is evident from Table 4.4. We should mention that this instance has 2 variables and no constraints in the first stage and 4 variables and 2 constraints in the second stage.

Table 4.4: Computational results for a small test instance

				MRP			MRPREG		
S	ϕ	Opt.	VCSP	Ite.	Cuts	CPU	Ite.	Cuts	CPU
4	50	0.500	0.750	3	3	0.07	4	3	0.10
9	50	0.111	0.556	4	9	0.09	5	9	0.08
36	50	0.167	0.306	3	17	0.11	4	17	0.12
121	50	0.132	0.182	5	58	0.34	6	58	0.38
441	50	0.120	0.152	6	216	1.65	8	216	2.06

Acknowledgment

The author would like to thank Andreas Märkert for providing the EPS problem instances for computational testing.

Chapter 5

The Minimax Problem

In this chapter our starting point is a general optimization problem in which some parameters are not known with certainty. As we have seen in previous chapters of this thesis, a typical approach in stochastic programming is to represent the uncertain parameters in such a problem by random variables with a known probability distribution, and subsequently to minimize some appropriate objective function representing e.g. total expected cost with respect to this distribution. As we have repeatedly pointed out, however, the probability distribution of uncertain parameters will most frequently not be completely known. Therefore, a large amount of research has gone into establishing stability results for stochastic programming problems, justifying the approximation of the true probability distribution by some suitable estimate. In fact, such an approach has been followed in Chapters 2, 3, and 4 of this thesis. In the present chapter, on the other hand, we consider the alternative minimax approach to stochastic programming. This approach generally concerns the problem of minimizing the worst expected value of the objective function with respect to the set of all probability distributions that are consistent with the available information on the random data.

5.1 Problem Formulation

Assuming that an outcome of random parameters constitutes a vector in $\mathbb{R}^N$, we will once again be concerned with $\mathcal{P}(\mathbb{R}^N)$, the set of all Borel probability measures on $\mathbb{R}^N$. As pointed out above, a classical approach in stochastic programming is to assume that the distribution $\mu \in \mathcal{P}(\mathbb{R}^N)$ of random parameters is known, and then to consider a problem of the general form

$$\min_{x \in X} \mathcal{Q}(x, \mu), \tag{5.1.1}$$

where X is some closed, convex set, and the function $\mathcal{Q}(x, \mu)$ typically denotes total expected cost given the decision x and the distribution μ.

A major concern in the formulation of the stochastic programming model (5.1.1) is the fact that the distribution μ of random parameters will hardly ever be directly accessible. In practice, the only information on the probability distribution that is available will most often at best be an estimate based on statistical information on the random data. In this situation an alternative approach is the following. First, a class $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^N)$ of possible or conceivable distributions, consistent with the available information, is established. A typical example is the case when $\mathcal{A}$ is the set of all distributions satisfying a number of mathematical constraints, taking the form of upper and lower bounds on mean, variance and other moments, but in general this class may be defined by any information characterizing probability distributions. Given the set of conceivable distributions $\mathcal{A}$ the following problem is considered,

$$MMP(\mathcal{A}) \qquad \min_{x \in X} \Big\{ f(x) = \sup_{\mu \in \mathcal{A}} \mathcal{Q}(x, \mu) \Big\}. \tag{5.1.2}$$

Problem (5.1.2) is referred to as the minimax approach to stochastic programming. Starting with Žáčková [167], variations of this approach have previously been considered by authors such as e.g. Birge and Dulá [20], Birge and Wets [23], Breton and Hachem [25, 26], Dupačová [44, 45], Ermoliev, Gaivoronski, and Nedeva [48], Kall [66], and Shapiro and Kleywegt [147].

Only few practicable solution procedures for the minimax problem (5.1.2) have been elaborated, and the ones proposed rely on results such as the following. If $\mathcal{A}$ is the class of all probability measures with support in some compact set Ξ, satisfying a number of generalized moment conditions,

$$\int_{\Xi} g_i(\xi)\mu(d\xi) \leq \alpha_i, \quad i = 1, \dots, L,$$

where $g_i(\cdot), i = 1 \dots, L$, are bounded continuous functions on Ξ, then $\mathcal{A}$ is a compact, convex set, and its extreme points are discrete measures with finite supports of at most $L+1$ points (cf. Karr [68, Theorem 2.1], see also Kempermann [69]). Thus, in this case, attention can be restricted to discrete measures in $\mathcal{A}$ having finite support of at most $L+1$ points. In Breton and Hachem [25, 26] this approach was employed to develop two alternative solution procedures, an extension of the progressive hedging algorithm (see Rockafellar and Wets [124]) and a bundle method, respectively. In fact, though, Breton and Hachem not only assume that all measures in $\mathcal{A}$ have finite support of at most $L+1$ points, but also that the support is known and identical for all $\mu \in \mathcal{A}$. The same approach is followed by Takriti and Ahmed [152] who consider a two-stage stochastic minimax problem arising in electricity trading, and propose a cut-and-branch procedure to solve the problem. In Ermoliev, Gaivoronski, and Nedeva [48], on the other hand, generalized linear programming techniques are employed to determine the mass points and corresponding probabilities of a "worst" distribution in $\mathcal{A}$ with $L+1$ mass points. These techniques are combined with a projected quasi subgradient approach to determine an optimal solution of the minimax problem. Situations may occur, however, in which the set $\mathcal{A}$ does not fit into the above-mentioned framework, so that the restriction of attention to measures with finite support is not immediately valid. Also, even if attention can be restricted to measures in $\mathcal{A}$ with finite support, the inner maximization in (5.1.2), taking the form e.g. of a generalized moment problem, may still be intractable. In such situations, stability analysis of the minimax problem, when the set of conceivable measures is subjected to perturbations, becomes relevant. Thus, the results presented in Section 5.2 below justify in particular the assumption that was implicitly made by Breton and Hachem [25, 26] and by Takriti and Ahmed [152], restricting attention to discrete measures with support in some known finite set.

5.1.1 Two-Stage Recourse Models

Throughout the structural analysis of the minimax problem presented in the following section, we will consider a general objective function $\mathcal{Q}$ as in (5.1.2). We will, however, be particularly concerned with application of the results to different classes of two-stage stochastic recourse models. From previous chapters of this thesis we recall that such models are based on the assumption that an alternating process of decisions and observations of random data is appropriate. More specifically, it is assumed that some decisions must be taken in a first stage in which only distributional information on the uncertainties is available. The outcome of random parameters is subsequently observed and some *recourse* actions may be taken in a second stage. A classical approach to this class of problems, followed in Chapters 2 and 3, is to minimize the sum of first-stage cost and expected second-stage, i.e. to let

$$\mathcal{Q}(x,\mu) = cx + \int_{\mathbb{R}^N} \Phi(x,\xi)\mu(d\xi), \tag{5.1.3}$$

where the random vector $\xi \in \mathbb{R}^N$ is constituted by the components of a tuple (q, h, T), and the second-stage value function Φ is given accordingly by

$$\Phi(x,\xi) = \min\bigl\{q(\xi)y \mid Wy \geq h(\xi) + T(\xi)x,\ y \in Y\bigr\}. \tag{5.1.4}$$

Here $c \in \mathbb{R}^{n_1}$ is a known vector, and W is a known rational matrix of size $m_2 \times n_2$, referred to as the recourse matrix. The second-stage cost $q : \mathbb{R}^N \mapsto \mathbb{R}^{n_2}$, the second-stage right-hand side $h : \mathbb{R}^N \mapsto \mathbb{R}^{m_2}$, and the technology matrix $T : \mathbb{R}^N \mapsto \mathbb{R}^{m_2 \times n_1}$, on the other hand, are represented as mappings picking out the appropriate components of the random vector ξ. Finally, the set $Y \subseteq \mathbb{R}^{n_2}$ may or may not contain integrality restrictions on some or all of the second-stage variables.

Remark 5.1.1. Even though we will primarily be concerned with the classical expected recourse function (5.1.3), we note that the results presented in this chapter may also be applied in connection with the non-classical approach followed in Chapter 4, where the probability of total cost exceeding some prescribed threshold value ϕ was minimized, i.e. we considered the probability-based recourse function

$$\mathcal{Q}(x,\mu) = \mu\bigl(\bigl\{\xi \in \mathbb{R}^N \mid cx + \Phi(x,\xi) > \phi\bigr\}\bigr), \tag{5.1.5}$$

where the second-stage value function is still defined by (5.1.4).

Remark 5.1.2. Under fairly general assumptions, the function $\mathcal{Q}(\cdot,\mu)$, defined by either (5.1.3) or (5.1.5), is a real-valued, lower semicontinuous functions on $\mathbb{R}^{n_1}$ for any $\mu \in \mathcal{P}(\mathbb{R}^N)$, cf. e.g. Theorem 3.2.1 on page 31 and Proposition 4.2.1 on page 50. Moreover, it is easily seen that the supremum of a family of lower semicontinuous functions is again a lower semicontinuous function (cf. the proof of Proposition 5.3.4 on page 78). Thus, when $\mathcal{Q}$ is defined by either (5.1.3) or (5.1.5), mild assumptions ensure that the minimax problem (5.1.2) is well-defined in the sense that one minimizes a lower semicontinuous function over a closed convex set, and hence the optimal value is actually attained, provided that the problem is feasible and bounded.

The reason for focusing on two-stage stochastic recourse problems here, is partly the fact that much research has gone into stability properties of optimal solutions of these problems, when the underlying probability distribution μ varies in some subset of $\mathcal{P}(\mathbb{R}^N)$, cf. the discussion in Sections 2.2, 3.2, and 4.2. In particular, qualitative and quantitative continuity properties of the particular recourse function $\mathcal{Q}$ lead to continuity results for the corresponding optimal value function and solution set mapping as functions of μ. Now, it turns out that the continuity properties of the recourse functions discussed in previous chapters are in fact also sufficient to provide similar stability results for the corresponding minimax problem (5.1.2), when the set $\mathcal{A}$ is subjected to perturbations. Furthermore, as already pointed out, these stability results for the minimax problem justify in particular the approach of restricting attention to probability measures with support in some known finite set. For applications of the minimax approach in the two-stage recourse setting, this allows us to develop straightforward extensions of the well-known solution procedures discussed in previous chapters to solve the corresponding minimax problems. Still, we note that the stability results presented here are applicable for any other class of problems for which similar continuity properties can be established.

5.2 Stability

In this section we establish qualitative and quantitative stability results for the minimax problem (5.1.2) when the set of conceivable distributions $\mathcal{A}$ is subjected to perturbations. To facilitate the analysis, we once again endow the set of all Borel probability measures on $\mathbb{R}^N$ with the notion of weak convergence (see Appendix A). The qualitative stability results will be based on the assumed joint continuity of $\mathcal{Q}$ with respect to x and μ, whereas the quantitative results rely on quantitative continuity properties of $\mathcal{Q}(x,\cdot)$ for $x \in X$. Sufficient conditions for these properties to hold when $\mathcal{Q}$ is defined by either (5.1.3) or (5.1.5) were presented in Sections 2.2.2, 3.2.1, and 4.2.2, respectively.

The stability results presented next take the form of continuity properties of the optimal value function and the solution set mapping as functions of $\mathcal{A}$. Since the problem is non-convex in several important cases, such as e.g. mixed-integer recourse models, we will include local optimizers in the analysis. To this end, we define for any non-empty open set $V \subseteq \mathbb{R}^{n_1}$, a localized version of the optimal value function,

$$\varphi_V(\mathcal{A}) = \min_{x \in X \cap \operatorname{cl} V} \sup_{\mu \in \mathcal{A}} \mathcal{Q}(x,\mu),$$

and a localized version of the solution set mapping,

$$\Psi_V(\mathcal{A}) = \big\{x \in X \cap \operatorname{cl} V \mid \sup_{\mu \in \mathcal{A}} \mathcal{Q}(x,\mu) = \varphi_V(\mathcal{A})\big\},$$

where $\operatorname{cl} V$ denotes the closure of V. Moreover, we will adopt the notion of a complete local minimizing (CLM) set, originally introduced by Robinson [121] and Klatte [73]. Given some set $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^N)$ of conceivable distributions, a set $M \subseteq \mathbb{R}^{n_1}$ is called a CLM set for $MMP(\mathcal{A})$ if there exists some open set $V \subseteq \mathbb{R}^{n_1}$ such that $M = \Psi_V(\mathcal{A})$ and $M \subseteq V$. Clearly, the set of global minimizers of problem (5.1.2), as well as any set of strict local minimizers, is a CLM set. Moreover, the notion of CLM sets precludes pathologies arising when $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^N)$ and $V \subseteq \mathbb{R}^{n_1}$ are such that $\Psi_V(\mathcal{A})$ is a set of local minimizers for problem (5.1.2), while even the slightest perturbation of $\mathcal{A}$ leads to local solution sets with respect to V, which do not contain any local minimizers of problem (5.1.2). (See e.g. Example 3.2.1 on page 34.)

Proposition 5.2.1. *Let $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^N)$ and let $\{\mathcal{A}_n\}_{n\geq 1}$ be a sequence of sets of Borel probability measures on $\mathbb{R}^N$ converging to $\mathcal{A}$ in the sense that*

(i) $\cup_{n\geq 1}\mathcal{A}_n \subseteq \mathcal{A}$ *and for all* $n \geq 1$ *we have* $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$;

(ii) for all $\mu \in \mathcal{A}$ *there exists a sequence* $\{\mu_n\}_{n\geq 1}$ *such that* $\mu_n \in \mathcal{A}_n$ *for all* $n \geq 1$ *and* $\mu_n \xrightarrow{w} \mu$.

Also, let $V \subseteq \mathbb{R}^{n_1}$ be some bounded open set such that $\Psi_V(\mathcal{A})$ is a CLM set for $MMP(\mathcal{A})$ with respect to V. If the function $\mathcal{Q}$ satisfies that

(iii) $\mathcal{Q} : \mathbb{R}^{n_1} \times \mathcal{P}(\mathbb{R}^N) \mapsto \mathbb{R}$ *is continuous at* (x, μ) *for all* $x \in X$ *and* $\mu \in \mathcal{A} \setminus \cup_{n\geq 1}\mathcal{A}_n$;

(iv) $\mathcal{Q}(\cdot, \mu)$ *is a lower semicontinuous function on* $\mathbb{R}^{n_1}$ *for all* $\mu \in \mathcal{A}$;

then it holds that

(a) $\varphi_V(\mathcal{A}_n) \to \varphi_V(\mathcal{A})$ *as* $n \to \infty$;

(b) $\sup_{x\in\Psi_V(\mathcal{A}_n)} \operatorname{dist}(x, \Psi_V(\mathcal{A})) \to 0$ *as* $n \to \infty$;

(c) there exists $N \geq 1$ *such that* $\Psi_V(\mathcal{A}_n)$ *is a CLM set for* $MMP(\mathcal{A}_n)$ *with respect to* V *for all* $n \geq N$.

Proof. Define for all $x \in X \cap \operatorname{cl} V$ and $n \geq 1$,

$$f(x) = \sup_{\mu\in\mathcal{A}} \mathcal{Q}(x, \mu) \quad \text{and} \quad f_n(x) = \sup_{\mu\in\mathcal{A}_n} \mathcal{Q}(x, \mu).$$

Now let $x^* \in \Psi_V(\mathcal{A})$ so that we have $\varphi_V(\mathcal{A}) = f(x^*)$, and for $n \geq 1$ let $x_n^* \in \Psi_V(\mathcal{A}_n)$ so that we have $\varphi_V(\mathcal{A}_n) = f_n(x_n^*)$.

The assumption that $\cup_{n\geq 1}\mathcal{A}_n \subseteq \mathcal{A}$ implies that for all $x \in X \cap \operatorname{cl} V$ and $n \geq 1$ we have $f_n(x) \leq f(x)$ so that $\limsup_{n\to\infty} \varphi_V(\mathcal{A}_n) \leq \varphi_V(\mathcal{A})$. To prove (a) by contradiction we assume that $\liminf_{n\to\infty} \varphi_V(\mathcal{A}_n) < \varphi_V(\mathcal{A}) - \epsilon$ for some $\epsilon > 0$. As a consequence of this assumption, there exists some infinite subset $\mathbb{K}_1 \subseteq \mathbb{N}$ such that $\varphi_V(\mathcal{A}_n) < \varphi_V(\mathcal{A}) - \epsilon$ for all $n \in \mathbb{K}_1$, and hence $\mathcal{Q}(x_n^*, \mu_n) < \varphi_V(\mathcal{A}) - \epsilon$ for all $\mu_n \in \mathcal{A}_n$ and all $n \in \mathbb{K}_1$. Now, compactness of $X \cap \operatorname{cl} V$ implies the existence of some infinite subset $\mathbb{K}_2 \subseteq \mathbb{K}_1$ such that the subsequence $\{x_n^*\}_{n\in\mathbb{K}_2}$ converges to some $\bar{x} \in X \cap \operatorname{cl} V$. Also, for any $\mu \in \mathcal{A}$ we can select a sequence $\{\mu_n\}_{n\in\mathbb{K}_2}$ converging weakly to μ such that $\mu_n \in \mathcal{A}_n$ for all $n \in \mathbb{K}_2$. Hence, for any $\mu \in \mathcal{A} \setminus \cup_{n\geq 1}\mathcal{A}_n$ we see by joint continuity of $\mathcal{Q}$ at $(\bar{x}, \mu)$ that

$$\mathcal{Q}(\bar{x}, \mu) = \lim_{\substack{n\to\infty \\ n\in\mathbb{K}_2}} \mathcal{Q}(x_n^*, \mu_n) < \varphi_V(\mathcal{A}) - \epsilon.$$

Finally, for any $\mu \in \cup_{n\geq 1}\mathcal{A}_n$ the condition $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$ implies that there exists $N \geq 1$ such that $\mu \in \mathcal{A}_n$ for all $n \geq N$. Hence, for any $\mu \in \cup_{n\geq 1}\mathcal{A}_n$ we see by lower semicontinuity of $\mathcal{Q}(\cdot, \mu)$ that we have

$$\mathcal{Q}(\bar{x}, \mu) \leq \liminf_{\substack{n\to\infty \\ n\in\mathbb{K}_2}} \mathcal{Q}(x_n^*, \mu) < \varphi_V(\mathcal{A}) - \epsilon.$$

Thus, for all $\mu \in \mathcal{A}$ we have $\mathcal{Q}(\bar{x}, \mu) < \varphi_V(\mathcal{A}) - \epsilon$ and hence $f(\bar{x}) < \varphi_V(\mathcal{A})$, a contradiction. This proves part (a).

To prove part (b), we let $\bar{x}$ be an accumulation point of the sequence $\{x_n^*\}_{n\geq 1}$, i.e. for some infinite subset $\mathbb{K}_3 \subseteq \mathbb{N}$ the sequence $\{x_n^*\}_{n\in\mathbb{K}_3}$ converges to $\bar{x}$. Let $\epsilon > 0$ be given and let $\bar{\mu} \in \mathcal{A}$ be such that $\mathcal{Q}(\bar{x}, \bar{\mu}) > f(\bar{x}) - \epsilon$. As above, whether $\bar{\mu} \in \mathcal{A} \setminus \cup_{n\geq 1}\mathcal{A}_n$ or $\bar{\mu} \in \cup_{n\geq 1}\mathcal{A}_n$, we can establish a sequence $\{\bar{\mu}_n\}_{n\in\mathbb{K}_3}$ of probability measures such that

$$\liminf_{\substack{n\to\infty \\ n\in\mathbb{K}_3}} \mathcal{Q}(x_n^*, \bar{\mu}_n) \geq \mathcal{Q}(\bar{x}, \bar{\mu}) > f(\bar{x}) - \epsilon$$

and $\bar{\mu}_n \in \mathcal{A}_n$ for all $n \in \mathbb{K}_3$. On the other hand, for any $n \in \mathbb{K}_3$ the definition of $f_n(\cdot)$ implies $f_n(x_n^*) \geq \mathcal{Q}(x_n^*, \bar{\mu}_n)$ and from part (a) we have

$$\lim_{n\to\infty} f_n(x_n^*) = \lim_{n\to\infty} \varphi_V(\mathcal{A}_n) = \varphi_V(\mathcal{A}).$$

Thus $f(\bar{x}) < \varphi_V(\mathcal{A}) + \epsilon$ and since ϵ was arbitrary we must have $f(\bar{x}) \leq \varphi_V(\mathcal{A})$ so that $\bar{x} \in \Psi(\mathcal{A})$. This proves part (b).

To prove part (c) we note that $\Psi_V(\mathcal{A}) \subseteq V$ where $\Psi_V(\mathcal{A})$ is closed and V is open and bounded. Therefore, for some $\delta > 0$ it must hold that $\mathrm{dist}(x, \Psi_V(\mathcal{A})) < \delta$ implies $x \in V$. Since non-emptiness of $\Psi_V(\mathcal{A}_n)$ for $n \geq 1$ follows from non-emptiness of $\Psi_V(\mathcal{A})$, boundedness of V, and lower semicontinuity of f_n, this completes the proof. □

Remark 5.2.1. Note that the assumption of joint continuity of $\mathcal{Q}$ at (x, μ) for all $x \in X$ and $\mu \in \mathcal{A} \setminus \cup_{n\geq 1}\mathcal{A}_n$ may in fact by relaxed to lower semicontinuity. However, non of the joint continuity results for the recourse function of two-stage stochastic recourse problems, presented in previous chapters, are restricted to lower semicontinuity, and hence we chose the formulation of Proposition 5.2.1. Also note that we only require joint continuity of $\mathcal{Q}$ for $\mu \in \mathcal{A} \setminus \cup_{n\geq 1}\mathcal{A}_n$ which is fortunate because the set $\cup_{n\geq 1}\mathcal{A}_n$ will typically contain all of the discrete measures in $\mathcal{A}$, and for these measures we cannot in general expect joint continuity, cf. e.g. the results concerning the expected recourse function of a two-stage stochastic program with mixed-integer recourse presented in Section 3.2.1.

Remark 5.2.2. The conclusion (b) of Proposition 5.2.1 implies that the set-valued mapping $\Psi_V(\cdot)$ is Hausdorff upper semicontinuous at $\mathcal{A}$ with respect to the notion of convergence of sets defined in the proposition cf. e.g. Deutsch, Pollul, and Singer [41, Lemma 1]. Furthermore, since the set $\Psi_V(\mathcal{A})$ is obviously compact the conclusion (b) actually implies Berge upper semicontinuity of $\Psi_V(\cdot)$ at $\mathcal{A}$ cf. the same lemma.

Remark 5.2.3. Proposition 5.2.1 applies for example in the situation when $\mathcal{A}$ is an infinite set of probability measures, rendering numerical solution of the inner maximization problem in (5.1.2) intractable. In this case Proposition 5.2.1 justifies the approach of approximating $\mathcal{A}$ by smaller subsets which are easier handled computationally. In particular, if $\mathcal{A}$ is the set of all Borel probability measures with support in some set Ξ, satisfying a number of generalized moment conditions, a possible approach is to restrict attention to those measures in $\mathcal{A}$ having support in some known finite set $\Xi' \subseteq \Xi$. In this way the inner maximization problem is simplified from a generalized moment problem to an ordinary linear program. Moreover, as a consequence of Proposition 5.2.1 we see that the optimal solution of such a simplified problem converges to the true optimal solution as the approximation of Ξ provided by Ξ' is progressively improved.

Clearly, different kinds of approximations of $\mathcal{A}$ than the one expressed in Proposition 5.2.1 may be more desirable in some situations. Consider for example the situation when $\mathcal{A}$ is constituted by a family of specific distributions, rendering the inner maximization problem in (5.1.2) intractable, e.g. $\mathcal{A}$ being a set of distributions that are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^N$. In this case, we do not want the approximations of $\mathcal{A}$ to be subsets thereof, but rather to be sets of probability measures approximating each of the individual measures in $\mathcal{A}$. We now establish a stability result that may be applied to justify such an approach. Before we proceed, though, we need to recall the following definition which may be found in e.g. Billingsley [18, Chapter 6]).

Definition 5.2.1. A set of probability measures $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^N)$ is said to be relatively compact if every sequence of measures from $\mathcal{A}$ contains a weakly convergent subsequence.

Remark 5.2.4. According to Definition 5.2.1, a set of probability measures $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^N)$ is said to be relatively compact if for every sequence $\{\mu_n\}_{n\geq 1}$ of measures from $\mathcal{A}$ there exist some measure $\mu \in \mathcal{P}(\mathbb{R}^N)$ (not necessarily belonging to $\mathcal{A}$) and an infinite subset $\mathbb{K} \subseteq \mathbb{N}$ such that the sequence $\{\mu_n\}_{n\in\mathbb{K}}$ converges weakly to μ — in that case we will refer to μ as an accumulation point of the sequence $\{\mu_n\}_{n\geq 1}$.

Proposition 5.2.2. *Let $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^N)$ and let $\{\mathcal{A}_n\}_{n\geq 1}$ be a sequence of sets of Borel probability measures converging to $\mathcal{A}$ in the sense that*

(i) if μ is an accumulation point of some sequence $\{\mu_n\}_{n\geq 1}$ where $\mu_n \in \mathcal{A}_n$ for all $n \geq 1$ then $\mu \in \mathcal{A}$;

(ii) for all $\mu \in \mathcal{A}$ there exists a sequence $\{\mu_n\}_{n\geq 1}$ such that $\mu_n \in \mathcal{A}_n$ for all $n \geq 1$ and $\mu_n \xrightarrow{w} \mu$;

and assume furthermore that $\cup_{n\geq 1}\mathcal{A}_n$ is relatively compact. Also, let $V \subseteq \mathbb{R}^{n_1}$ be some bounded open set such that $\Psi_V(\mathcal{A})$ is a CLM set for $MMP(\mathcal{A})$ with respect to V. If the function $\mathcal{Q}$ satisfies that

(iii) $\mathcal{Q} : \mathbb{R}^{n_1} \times \mathcal{P}(\mathbb{R}^N) \mapsto \mathbb{R}$ is continuous at (x, μ) for all $x \in X$ and $\mu \in \mathcal{A}$;

then it holds that

(a) $\varphi_V(\mathcal{A}_n) \to \varphi_V(\mathcal{A})$ as $n \to \infty$;

(b) $\sup_{x\in\Psi_V(\mathcal{A}_n)} \operatorname{dist}(x, \Psi_V(\mathcal{A})) \to 0$ as $n \to \infty$;

(c) there exists $N \geq 1$ such that $\Psi_V(\mathcal{A}_n)$ is a CLM set for $MMP(\mathcal{A}_n)$ with respect to V for all $n \geq N$.

Proof. Define for all $x \in X \cap \operatorname{cl} V$ and $n \geq 1$

$$f(x) = \sup_{\mu\in\mathcal{A}} \mathcal{Q}(x, \mu) \quad \text{and} \quad f_n(x) = \sup_{\mu\in\mathcal{A}_n} \mathcal{Q}(x, \mu).$$

Now let $x^* \in \Psi_V(\mathcal{A})$ so that we have $\varphi_V(\mathcal{A}) = f(x^*)$, and for $n \geq 1$ let $x_n^* \in \Psi_V(\mathcal{A}_n)$ so that we have $\varphi_V(\mathcal{A}_n) = f_n(x_n^*)$.

As in the proof of Proposition 5.2.1 we may show that $\liminf_{n\to\infty} \varphi_V(\mathcal{A}_n) \geq \varphi_V(\mathcal{A})$. Hence to prove part (a) by contradiction we assume that $\limsup_{n\to\infty} \varphi_V(\mathcal{A}_n) > \varphi_V(\mathcal{A})+\epsilon$

for some $\epsilon > 0$. Since we have $f_n(x_n^*) \le f_n(x^*)$ for all $n \ge 1$, the assumption implies $\limsup_{n\to\infty} f_n(x^*) > \varphi_V(\mathcal{A}) + \epsilon$, and hence there exist some infinite subset $\mathbb{K}_1 \in \mathbb{N}$ and measures $\mu_n \in \mathcal{A}_n$ for $n \in \mathbb{K}_1$ such that $\mathcal{Q}(x^*, \mu_n) > \varphi_V(\mathcal{A}) + \epsilon$ for all $n \in \mathbb{K}_1$. Also, by the assumption that $\cup_{n\ge1}\mathcal{A}_n$ is relatively compact, there exists some infinite subset $\mathbb{K}_2 \subseteq \mathbb{K}_1$ such that the sequence $\{\mu_n\}_{n\in\mathbb{K}_2}$ converges weakly to some μ. Now, according to assumption (i) we have $\mu \in \mathcal{A}$, and by joint continuity of $\mathcal{Q}$ at (x^*, μ) we see that

$$\mathcal{Q}(x^*, \mu) = \lim_{\substack{n\to\infty \\ n\in\mathbb{K}_2}} \mathcal{Q}(x^*, \mu_n) > \varphi_V(\mathcal{A}), \tag{5.2.1}$$

a contradiction. This proves part (a). The proofs of part (b) and (c) are similar to those in Proposition 5.2.1. □

Remark 5.2.5. The convergence of the sequence $\{\mathcal{A}_n\}_{n\ge1}$ to $\mathcal{A}$ defined in Proposition 5.2.2 is in fact the Kuratowski-Painlevé convergence of sets. (See e.g. Kuratowski [80].)

Remark 5.2.6. According to Prohorov's Theorem, the set $\cup_{n\ge1}\mathcal{A}_n$ is relatively compact if it is tight; that is, if for any $\epsilon > 0$ there exists a compact set Ξ such that $\mu(\Xi) > 1-\epsilon$ for all $\mu \in \cup_{n\ge1}\mathcal{A}_n$. In particular, if there exists some compact set $\Xi \subseteq \mathbb{R}^N$ such that all measures in $\cup_{n\ge1}\mathcal{A}_n$ have support in Ξ, then $\cup_{n\ge1}\mathcal{A}_n$ is relatively compact. (See e.g. Billingsley [18, Chapter 6].)

Remark 5.2.7. In line with Remark 5.2.2 we note that conclusion (b) of Proposition 5.2.2 implies Hausdorff upper semicontinuity as well as Berge upper semicontinuity of $\Psi_V(\cdot)$ at $\mathcal{A}$ with respect to the Kuratowski-Painlevé convergence of sets defined in the proposition. Note, however, that Remark 5.2.1 does not hold true for Proposition 5.2.2. More specifically, the continuity assumption for $\mathcal{Q}$ cannot be relaxed to lower semicontinuity since upper semicontinuity is necessary for the proof of part (a) cf. (5.2.1).

Remark 5.2.8. Let $\mathcal{A} = \{\mu^1, \ldots, \mu^k\} \subseteq \mathcal{P}(\mathbb{R}^N)$ and let $\{\mathcal{A}_n\}_{n\ge1}$ be a sequence of sets of Borel probability measures converging to $\mathcal{A}$ in the sense that for $n \ge 1$ we have $\mathcal{A}_n = \{\mu_n^1, \ldots, \mu_n^k\}$ where $\mu_n^j \xrightarrow{w} \mu^j$ for $j = 1, \ldots, k$. It is easily seen that this is in fact a special case of Proposition 5.2.2. Thus, any accumulation point of a sequence of measures from $\cup_{n\ge1}\mathcal{A}_n$ must necessarily be one of the measures in $\mathcal{A}$, and clearly any sequence of measures from $\cup_{n\ge1}\mathcal{A}_n$ has at least one weakly convergent subsequence.

Next, we turn to quantitative stability results for the minimax problem (5.1.2). To this end we assume that D is some distance defined on the set $\mathcal{P}(\mathbb{R}^N)$. Also, for $\mu \in \mathcal{P}(\mathbb{R}^N)$ and $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^N)$ we let $D(\mu, \mathcal{A}) = \inf_{\nu\in\mathcal{A}} D(\mu, \nu)$ and introduce a Hausdorff-like distance between sets of probability measures, defined for $\mathcal{A}, \mathcal{B} \subseteq \mathcal{P}(\mathbb{R}^N)$ by

$$D_H(\mathcal{A}, \mathcal{B}) = \max\Big\{\sup_{\mu\in\mathcal{A}} D(\mu, \mathcal{B}) \, , \, \sup_{\mu\in\mathcal{B}} D(\mu, \mathcal{A})\Big\}$$

Proposition 5.2.3. *Let $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^N)$ and let $V \subseteq \mathbb{R}^{n_1}$ be some bounded open set such that $\Psi_V(\mathcal{A})$ is a CLM set for $MMP(\mathcal{A})$ with respect to V. If there exist constants $L, p, \delta > 0$ such that $|\mathcal{Q}(x, \mu) - \mathcal{Q}(x, \nu)| \le L \cdot D(\mu, \nu)^p$ whenever $x \in X$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$ with $D(\mu, \nu) < \delta$, then*

$$\big|\varphi_V(\mathcal{A}) - \varphi_V(\mathcal{B})\big| \le L \cdot D_H(\mathcal{A}, \mathcal{B})^p$$

whenever $\mathcal{B} \subseteq \mathcal{P}(\mathbb{R}^N)$ with $D_H(\mathcal{A}, \mathcal{B}) < \delta$.

Proof. Let $\mathcal{B} \subseteq \mathcal{P}(\mathbb{R}^N)$ with $D_H(\mathcal{A}, \mathcal{B}) < \delta$ and define for all $x \in X \cap \operatorname{cl} V$,

$$f(x) = \sup_{\mu \in \mathcal{A}} \mathcal{Q}(x, \mu) \quad \text{and} \quad g(x) = \sup_{\mu \in \mathcal{B}} \mathcal{Q}(x, \mu).$$

Now let $x_A^* \in \Psi_V(\mathcal{A})$ and $x_B^* \in \Psi_V(\mathcal{B})$ so that $\varphi_V(\mathcal{A}) = f(x_A^*)$ and $\varphi_V(\mathcal{B}) = g(x_B^*)$. Obviously, we have

$$\varphi_V(\mathcal{A}) - \varphi_V(\mathcal{B}) \leq f(x_B^*) - g(x_B^*).$$

Let $0 < \epsilon < \delta - D_H(\mathcal{A}, \mathcal{B})$ be given, and let $\bar{\mu} \in \mathcal{A}$ be such that $\mathcal{Q}(x_B^*, \bar{\mu}) > f(x_B^*) - \epsilon$. Also, let $\bar{\nu} \in \mathcal{B}$ be such that $D(\bar{\mu}, \bar{\nu}) \leq D_H(\mathcal{A}, \mathcal{B}) + \epsilon < \delta$ and $\mathcal{Q}(x_B^*, \bar{\nu}) \leq g(x_B^*)$. Then,

$$\begin{aligned} f(x_B^*) - g(x_B^*) &< \mathcal{Q}(x_B^*, \bar{\mu}) - \mathcal{Q}(x_B^*, \bar{\nu}) + \mathcal{Q}(x_B^*, \bar{\nu}) - g(x_B^*) + \epsilon \\ &\leq \big|\mathcal{Q}(x_B^*, \bar{\mu}) - \mathcal{Q}(x_B^*, \bar{\nu})\big| + \epsilon \\ &\leq L \cdot \big(D_H(\mathcal{A}, \mathcal{B}) + \epsilon\big)^p + \epsilon. \end{aligned}$$

Since ϵ was arbitrary we get

$$\varphi_V(\mathcal{A}) - \varphi_V(\mathcal{B}) \leq L \cdot D_H(\mathcal{A}, \mathcal{B})^p,$$

and in the exact same way we may show that

$$\varphi_V(\mathcal{B}) - \varphi_V(\mathcal{A}) \leq L \cdot D_H(\mathcal{A}, \mathcal{B})^p.$$

Thus we have

$$\big|\varphi_V(\mathcal{A}) - \varphi_V(\mathcal{B})\big| \leq L \cdot D_H(\mathcal{A}, \mathcal{B})^p,$$

and the proof is complete. □

Remark 5.2.9. Assuming that the set $\mathcal{A}$ is compact, it may be seen that if the distance D is such that it metrizes the topology of weak convergence, then the result in Proposition 5.2.3 quantifies those in Proposition 5.2.1 (a) and Proposition 5.2.2 (a). First, let $\{\mathcal{A}_n\}_{n\geq 1}$ be a sequence of sets of probability measures converging to $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^N)$ in the sense of Proposition 5.2.1. Then $\cup_{n\geq 1}\mathcal{A}_n \subseteq \mathcal{A}$ implies $\sup_{\mu_n \in \mathcal{A}_n} D(\mu_n, \mathcal{A}) = 0$ for $n \geq 1$. Also, for any $\mu \in \mathcal{A}$ there exists a sequence $\{\mu_n\}_{n\geq 1}$ such that $\mu_n \in \mathcal{A}_n$ for all $n \geq 1$ and $\mu_n \xrightarrow{w} \mu$, and hence $D(\mu, \mathcal{A}_n) \to 0$ as $n \to \infty$. To see that $\sup_{\mu \in \mathcal{A}} D(\mu, \mathcal{A}_n) \to 0$ as $n \to \infty$ assume on the contrary that there exist an $\epsilon > 0$ and an infinite subset $\mathbb{K}_1 \subseteq \mathbb{N}$ such that for all $n \in \mathbb{K}_1$ there exist $\mu_n \in \mathcal{A}$ with $D(\mu_n, \mathcal{A}_n) > \epsilon$. Now, assuming that $\mathcal{A}$ is compact, there exist some $\mu \in \mathcal{A}$ and an infinite subset $\mathbb{K}_2 \subseteq \mathbb{K}_1$ such that the sequence $\{\mu_n\}_{n\in\mathbb{K}_2}$ converges weakly to μ. Since we obviously have for all $n \in \mathbb{K}_2$ that $\epsilon < D(\mu_n, \mathcal{A}_n) \leq D(\mu_n, \mu) + D(\mu, \mathcal{A}_n)$, we see that $D(\mu, \mathcal{A}_n) \not\to 0$ as $n \to \infty$, a contradiction. Therefore, we must have the desired result, $D_H(\mathcal{A}_n, \mathcal{A}) \to 0$ as $n \to \infty$. Next, assume that $\{\mathcal{B}_n\}_{n\geq 1}$ is a sequence of sets of probability measures converging to $\mathcal{B} \subseteq \mathcal{P}(\mathbb{R}^N)$ in the sense of Proposition 5.2.2 and such that $\cup_{n\geq 1}\mathcal{B}_n$ is relatively compact. Then we see in the same way as before that $\sup_{\mu \in \mathcal{B}} D(\mu, \mathcal{B}_n) \to 0$ as $n \to \infty$. Now, assume that $\sup_{\mu_n \in \mathcal{B}_n} D(\mu_n, \mathcal{B}) \not\to 0$ as $n \to \infty$. Then there exist $\epsilon > 0$, an infinite subset $\mathbb{K}_3 \subseteq \mathbb{N}$ and a sequence of measures $\{\mu_n\}_{n\in\mathbb{K}_3}$ with $\mu_n \in \mathcal{B}_n$ and $D(\mu_n, \mathcal{B}) > \epsilon$ for all $n \in \mathbb{K}_3$. Since $\cup_{n\geq 1}\mathcal{B}_n$ is relatively compact there exists some infinite subset $\mathbb{K}_4 \subseteq \mathbb{K}_3$ such that the sequence $\{\mu_n\}_{n\in\mathbb{K}_4}$ converges weakly to some μ which by assumption must belong to $\mathcal{B}$, a contradiction. Hence we see that $D_H(\mathcal{B}_n, \mathcal{B}) \to 0$ as $n \to \infty$.

Remark 5.2.10. Consider the formulation of a classical two-stage stochastic program with recourse, where the expected recourse function $\mathcal{Q}$ is given by (5.1.3)-(5.1.4). When the problem has linear recourse, i.e. $Y = \mathbb{R}_+^{n_2}$, basic assumptions guarantee that $\mathcal{Q}(\cdot, \mu)$ is a well-defined, real-valued, Lipschitzian, and convex function on $\mathbb{R}^{n_1}$ for any $\mu \in \mathcal{P}(\mathbb{R}^N)$ cf. Theorem 2.2.3 on page 18. Joint continuity of $\mathcal{Q}$ with respect to x and μ, on the other hand, typically requires some uniform integrability condition to be satisfied cf. Theorem 2.2.4 on page 18. Hence, conditions under which the qualitative stability results of Proposition 5.2.1 and Proposition 5.2.2 may be applied are well-known. Furthermore, to arrive at quantitative continuity results of $\mathcal{Q}(x, \cdot)$ for $x \in X$ of the form required in Proposition 5.2.3, the set $\mathcal{P}(\mathbb{R}^N)$ can be equipped for example with the bounded Lipschitz metric or an L_p-Wasserstein metric cf. Theorem 2.2.5 on page 19 and Theorem 2.2.6 on page 20, respectively.

Remark 5.2.11. Consider again the formulation of the expected recourse function $\mathcal{Q}$ given by (5.1.3)-(5.1.4), and assume now that the problem has (mixed-) integer recourse, i.e. that the set Y contains integrality restrictions on some or all of the second-stage variables. Structural properties of such problems have mainly been investigated for the case of fixed second-stage cost q in which case the expected recourse function $\mathcal{Q}(\cdot, \mu)$ is lower semicontinuous for any $\mu \in \mathcal{P}(\mathbb{R}^N)$ cf. Theorem 3.2.1 on page 31. To arrive at joint continuity of $\mathcal{Q}$ at some point (x, μ) in this setting, the above-mentioned uniform integrability condition must be combined with some assumption guaranteeing that the set of those $\xi \in \mathbb{R}^N$, for which the second-stage value function (5.1.4) is discontinuous at (x, ξ), has μ-measure zero cf. Theorem 3.2.4 on page 32. Finally, for problems with fixed technology matrix T the results may be quantified using a certain variational distance cf. Theorem 3.2.5 on page 33. Hence, also in this case, conditions under which the stability results presented above may be applied are well-known.

Remark 5.2.12. In Chapter 4 we considered the so-called minimum risk problem where the recourse function is defined by (5.1.5). It was shown that this formulation is in fact equivalent to a classical two-stage stochastic program with mixed-integer recourse and hence is a special case of (5.1.3). Moreover, it was shown that the above-mentioned continuity properties remain valid, even under simplified assumptions, cf. Proposition 4.2.1, Proposition 4.2.3, and Proposition 4.2.4 on page 50-53.

5.3 Solution Procedures

In this section we elaborate solution procedures for the minimax problem in the setting of two-stage stochastic recourse models, considering the linear recourse case as well as the integer recourse case. Hence, the recourse function is given by

$$\mathcal{Q}(x, \mu) = cx + \int_{\mathbb{R}^N} \Phi(x, \xi)\mu(d\xi), \tag{5.3.1}$$

where the random vector $\xi \in \mathbb{R}^N$ corresponds to an outcome of the second-stage cost $q(\xi)$, the second-stage right-hand side $h(\xi)$, and the technology matrix $T(\xi)$, and the second-stage value function Φ is defined accordingly as the value function of a linear program or an integer program. We will make the following assumption.

(D1) The set of conceivable distributions $\mathcal{A}$ is defined as the set of all probability measures μ with support in some finite set $\Xi = \{\xi^1, \ldots, \xi^S\}$, satisfying a number of generalized moment conditions,

$$\int_{\Xi} g_i(\xi)\mu(d\xi) \leq \alpha_i, \quad i = 1, \ldots, L.$$

According to (D1) attention is restricted to a finite number of scenarios, each scenario $s \in \{1, \ldots, S\}$ corresponding to an outcome of random parameters $(q(\xi^s), h(\xi^s), T(\xi^s))$. For ease of notation we will refer to such an outcome simply by (q^s, h^s, T^s).

Remark 5.3.1. Note that assumption (D1) is justified by the stability results established in the previous section, since the optimal solution of a minimax problem, employing a more general definition of the set of conceivable distributions, may be approximated to any given accuracy by solutions of minimax problems, employing only sets of probability measures with support in known finite sets (cf. Remark 5.2.3, Remark 5.2.10, and Remark 5.2.11).

Remark 5.3.2. Since the solution procedures presented below are all modifications of well-known algorithms, their efficacy is immediate from the computational testing of these original procedures and hence we do not report results of any computational experiments.

5.3.1 Two-Stage Linear Recourse Models

In this section, we consider the minimax problem (5.1.2) in the setting of a two-stage stochastic program with linear recourse, i.e. the second-stage value function is defined by

$$\Phi(x, \xi^s) = \min\{q^s y \mid Wy \geq h^s + T^s x,\ y \in \mathbb{R}_+^{n_2}\}, \quad s = 1, \ldots, S. \tag{5.3.2}$$

We elaborate a solution procedure for the problem under the following assumptions.

(D2) For all $t \in \mathbb{R}^{m_2}$ there exists $y \in \mathbb{R}_+^{n_2}$ such that $Wy \geq t$.

(D3) For all $s \in \{1, \ldots, S\}$ there exists $u \in \mathbb{R}_+^{m_2}$ such that $uW \leq q^s$.

Here (D2) is the assumption of complete recourse, ensuring feasibility of the second-stage problem for any right-hand side $t \in \mathbb{R}^{m_2}$, whereas (D3) is the assumption of dual feasibility, employed to ensure boundedness of the second-stage problems.

Remark 5.3.3. As pointed out also in previous chapters, we note that for practical purposes it is often sufficient to replace (D2) by the weaker assumption of relatively complete recourse, ensuring feasibility of the second-stage problem only for those right-hand sides that may actually occur, i.e. for all $x \in X$ and all $s \in \{1, \ldots, S\}$ there exists $y \in \mathbb{R}_+^{n_2}$ such that $Wy \geq h^s + T^s x$. Furthermore, we note that if relatively complete recourse is not inherent in the problem, it may be established by the inclusion of feasibility cuts cf. for example the discussion of the L-shaped algorithm in Section 2.3.1.

Employing assumption (D1) we may reformulate the minimax problem (5.1.2) in terms of the scenario probabilities $p_1, \ldots, p_S$ as follows

$$\min_{x \in X}\left\{ f(x) = cx + \max_{p \in \mathcal{P}} \sum_{s=1}^{S} p_s \Phi(x, \xi^s) \right\} \tag{5.3.3}$$

where

$$\mathcal{P} = \left\{ p \in \mathbb{R}^S_+ \;\middle|\; \sum_{s=1}^{S} p_s = 1,\; \sum_{s=1}^{S} p_s g_i(\xi^s) \le \alpha_i,\; i = 1\ldots, L \right\}. \tag{5.3.4}$$

Remark 5.3.4. Note that the inner maximization problem is now a linear programming problem over a bounded polyhedron and hence the maximum value is always attained.

The following proposition states the relevant structural properties of the function f.

Proposition 5.3.1. *Assume (D1)-(D3) and let f be defined by (5.3.2) and (5.3.3). Then f is a real-valued, piecewise linear, and convex function on X.*

Proof. It is well-known that for each $p \in \mathcal{P}$ the function $g_p(x) = cx + \sum_{s=1}^{S} p_s \Phi(x, \xi^s)$ is a real-valued, piecewise linear, and convex function on X, cf. e.g. Birge and Louveaux [22]. Noting that $\mathcal{P}$ is a bounded polyhedron, we see that f is the maximum of a finite number of real-valued, piecewise linear, and convex functions, corresponding to the extreme points of $\mathcal{P}$. □

The solution procedure that we will now propose is a modification of the L-shaped algorithm discussed in Section 2.3.1. It is based on the following reformulation of problem (5.3.3),

$$\min\ cx + \theta \tag{5.3.5a}$$

$$\text{s.t. } \theta \ge \sum_{s=1}^{S} p_s \Phi(x, \xi^s), \quad p \in \mathcal{P}, \tag{5.3.5b}$$

$$x \in X,\ \theta \in \mathbb{R}. \tag{5.3.5c}$$

As in L-shaped decomposition, the constraints of this problem may be replaced by linear inequalities referred to as optimality cuts. The algorithm progresses by sequentially solving a master problem and adding optimality cuts which are violated at the current solution. The master problem is initially obtained by removing the constraints (5.3.5b) from problem (5.3.5). Then, given a solution (x^ν, θ^ν) of the master problem in iteration ν, the second-stage problems are solved to obtain $w^{s,\nu} = \Phi(x^\nu, \xi^s) = \pi^{s,\nu}(h^s + T^s x^\nu)$ where $\pi^{s,\nu}$ are optimal dual solutions for $s = 1, \ldots, S$. Next, denoting by p^ν the optimal solution of the linear programming problem $w^\nu = \max_{p \in \mathcal{P}} \sum_{s=1}^{S} w^{s,\nu} p_s$, we obtain the following inequality that is valid for all $x \in X$ and binding at $x = x^\nu$,

$$\max_{p \in \mathcal{P}} \sum_{s=1}^{S} p_s \Phi(x, \xi^s) \ge \sum_{s=1}^{S} p_s^\nu \pi^{s,\nu}(h^s + T^s x).$$

In this way we see that if $\theta^\nu < w^\nu$, the current solution (x^ν, θ^ν) may be cut off by including the following constraint in the master problem,

$$\theta \ge \sum_{s=1}^{S} p_s^\nu \pi^{s,\nu}(h^s + T^s x).$$

Algorithm 5.1

Step 1 (*Initialization*) Let $K > 0$, set $\nu = 0$ and $\bar{z} = \infty$, and let the current master problem be $\min\{cx + \theta \mid x \in X,\ \theta \in \mathbb{R}\}$.

Step 2 (*Solve master problem*) Set $\nu = \nu + 1$. Solve the current master problem and let (x^ν, θ^ν) be an optimal solution vector if one exists; if the problem is unbounded, then let (x^ν, θ^ν) be a feasible solution with $cx^\nu + \theta^\nu < \bar{z} - K$.

Step 3 (*Termination*) If $cx^\nu + \theta^\nu = \bar{z}$, stop; the current solution is optimal.

Step 4 (*Solve subproblems*) For each $s \in \{1, \dots, S\}$, solve the second-stage problem to find $w^{s,\nu} = \Phi(x^\nu, \xi^s)$ and let $\pi^{s,\nu}$ be a corresponding optimal dual solution.

Step 5 (*Solve max-problem*) Solve the problem $w^\nu = \max_{p \in \mathcal{P}} \sum_{s=1}^S w^{s,\nu} p_s$ and let p^ν be an optimal solution. If $\theta^\nu < w^\nu$, add the cut $\theta \geq \sum_{s=1}^S p_s^\nu \pi^{s,\nu}(h^s + T^s x)$ to the master problem.

Step 6 (*Update bound*) Let $\bar{z} = \min\{\bar{z}, cx^\nu + w^\nu\}$. Go to Step 2.

It is easily seen that Algorithm 5.1 terminates in a finite number of iterations whenever a solution of the minimax problem exists.

Proposition 5.3.2. *Assume (D1)-(D3). If problem (5.3.5) is feasible and bounded, then Algorithm 5.1 terminates with an optimal solution in a finite number of iterations.*

Proof. Assume that the minimax problem has an optimal solution x^*. In any iteration ν of the algorithm we must have $cx^\nu + \theta^\nu \leq cx^* + \max_{p \in \mathcal{P}} \sum_{s=1}^S p_s \Phi(x^*, \xi^s)$, since the master problem is a relaxation of problem (5.3.5). As mentioned above, the current solution (x^ν, θ^ν) is cut off by an optimality cut whenever $\theta^\nu < \max_{p \in \mathcal{P}} \sum_{s=1}^S p_s \Phi(x^\nu, \xi^s)$. This can only happen a finite number of times since the number of optimality cuts is finite, cf. Proposition 5.3.1. Thus we will eventually have $\theta^\nu = \max_{p \in \mathcal{P}} \sum_{s=1}^S p_s \Phi(x^\nu, \xi^s)$, at which point the current solution is feasible for problem (5.3.5) and hence optimal. □

Remark 5.3.5. Takriti and Ahmed [152] consider a two-stage stochastic minimax problem arising in electricity trading. The problem has linear recourse but is complicated by the fact that first-stage variables are restricted to binaries. The authors propose a cutting-plane procedure to solve the problem, basically embedding the above procedure in a branch-and-cut scheme. The successful computational experiments reported in the paper, confirm our conjecture that the straightforward modification of the L-shaped algorithm is an efficient way to handle two-stage stochastic minimax problems with linear recourse.

Consider now the following alternative reformulation of problem (5.3.3),

$$
\begin{aligned}
&\min\ cx + \theta \\
&\text{s.t. } \theta \geq \sum_{s=1}^S p_s \sigma_s, \quad p \in \mathcal{P}, \\
&\qquad \sigma_s \geq \Phi(x, \xi^s), \quad s = 1, \dots, S, \\
&\qquad x \in X,\ \theta \in \mathbb{R},\ \sigma \in \mathbb{R}^S.
\end{aligned}
$$

This reformulation leads directly to the following multicut version of Algorithm 5.1.

Algorithm 5.2

Step 1 (*Initialization*) Let $K > 0$, set $\nu = 0$ and $\bar{z} = \infty$, and let the current master problem be $\min\{cx + \theta \mid x \in X,\ \theta \in \mathbb{R},\ \sigma \in \mathbb{R}^S\}$.

Step 2 (*Solve master problem*) Set $\nu = \nu + 1$. Solve the current master problem and let $(x^\nu, \theta^\nu, \sigma^\nu)$ be an optimal solution vector if one exists; if the problem is unbounded, then let $(x^\nu, \theta^\nu, \sigma^\nu)$ be a feasible solution with $cx^\nu + \theta^\nu < \bar{z} - K$.

Step 3 (*Termination*) If $cx^\nu + \theta^\nu = \bar{z}$, stop; the current solution is optimal.

Step 4 (*Solve subproblems*) For each $s \in \{1, \dots, S\}$, solve the second-stage problem to find $w^{s,\nu} = \Phi(x^\nu, \xi^s)$ and let $\pi^{s,\nu}$ be a corresponding optimal dual solution. If $\sigma_s^\nu < w^{s,\nu}$, add the cut $\sigma_s \geq \pi^{s,\nu}(h^s + T^s x)$ to the master problem.

Step 5 (*Solve max-problem*) Solve the problem $w^\nu = \max_{p \in \mathcal{P}} \sum_{s=1}^S w^{s,\nu} p_s$ and let p^ν be an optimal solution. If $\theta^\nu < \sum_{s=1}^S p_s^\nu \sigma_s^\nu$, add the cut $\theta \geq \sum_{s=1}^S p_s^\nu \sigma_s$ to the master problem.

Step 6 (*Update bound*) Let $\bar{z} = \min\{\bar{z}, cx^\nu + w^\nu\}$. Go to Step 2.

Proposition 5.3.3. *Assume (D1)-(D3). If problem (5.3.5) is feasible and bounded then Algorithm 5.2 terminates with an optimal solution in a finite number of iterations.*

Proof. The proof is similar to that of Proposition 5.3.2. □

Remark 5.3.6. The multicut approach of Algorithm 5.2 offers some computational advantages compared to the single-cut approach of Algorithm 5.1, since more detailed information is passed to the master problem in each iteration. The improved detailing, however, comes at the cost of an increased complexity of the master problem since the size of the problem grows quite rapidly. Also, because of their resemblance to the original L-shaped method and its multicut version, respectively, the algorithms presented above are bound to suffer from some of the same drawbacks. Apart from the growing size of the master problem, such drawbacks include the tendency for early iterations to oscillate heavily, causing slow convergence toward an optimal solution. In the case of L-shaped decomposition some of these drawbacks were circumvented by the regularized decomposition method discussed in Section 2.3.2. The idea is to introduce an incumbent solution a^ν and include a quadratic regularizing term of the form $\frac{\alpha}{2}\|x - a^\nu\|^2$ in the objective of the master problem. Clearly, a similar approach could be used for the algorithm presented here, but we will not go into the details of such an implementation.

Remark 5.3.7. In Chapter 4 we elaborated a solution procedure for the minimum risk problem, seeking the minimum value of the probability-based recourse function

$$\mathcal{Q}(x, \mu) = \mu\big(\{\xi \in \mathbb{R}^N \mid cx + \Phi(x, \xi) > \phi\}\big), \tag{5.3.6}$$

where ϕ is some given threshold value. This procedure is in many ways similar to the multicut version of the L-shaped algorithm. Hence we may modify Algorithm 5.2 in a similar way to obtain a solution procedure for the minimax problem (5.1.2) with $\mathcal{Q}$ defined by (5.3.6).

5.3.2 Two-Stage Integer Recourse Models

If second-stage variables are restricted to integer values, the solution procedures presented in the previous section break down, since the second-stage value function is no longer convex and piecewise linear, but in fact only lower semicontinuous. To solve the minimax problem (5.1.2) in this setting, we elaborate an extension of the branch-and-bound algorithm for two-stage stochastic programs with integer recourse, proposed by Ahmed, Tawarmalani, and Sahinidis [1] and discussed in Section 3.3.1. As in (5.3.1) the recourse function is defined as the sum of first-stage cost and expected second-stage cost, whereas the second-stage value function is now given by

$$\Phi(x, \xi^s) = \min\Big\{ q^s y \;\Big|\; Wy \geq h^s + T^s x,\ y \in \mathbb{Z}_+^{n_2} \Big\}, \quad s = 1, \dots, S. \tag{5.3.7}$$

The solution procedure is elaborated under the following assumptions.

(D2') For all $t \in \mathbb{R}^{m_2}$ there exists $y \in \mathbb{Z}_+^{n_2}$ such that $Wy \geq t$.

(D3) For all $s \in \{1, \dots, S\}$ there exists $u \in \mathbb{R}_+^{m_2}$ such that $uW \leq q^s$.

Note that (D2') is a natural extension of the complete recourse assumption for the integer recourse case, whereas the assumption on dual feasibility (D3) is unchanged.

Employing assumption (D1), we may once again reformulate the minimax problem (5.1.2) as

$$\min_{x \in X} \Big\{ f(x) = cx + \max_{p \in \mathcal{P}} \sum_{s=1}^{S} p_s \Phi(x, \xi^s) \Big\}, \tag{5.3.8}$$

where the set $\mathcal{P}$ is still defined by (5.3.4). According to the following proposition, problem (5.3.8) is well-defined in the sense that one minimizes a real-valued, lower semicontinuous function.

Proposition 5.3.4. *Assume (D1)-(D3), and let f be defined by (5.3.7) and (5.3.8). Then f is a real-valued, lower semicontinuous function on X.*

Proof. For each $p \in \mathcal{P}$ the function $g_p(x) = cx + \sum_{s=1}^{S} p_s \Phi(x, \xi^s)$ is a real-valued, lower semicontinuous function on X, cf. e.g. Nemhauser and Wolsey [97]. As in the proof of Proposition 5.3.1 we see that f is real-valued. Now, let $x \in X$ and let $\{x_n\}_{n \geq 1}$ be some sequence in X converging to x. Assuming that $\bar{p} \in \arg\max_{p \in \mathcal{P}} \{g_p(x)\}$, we have

$$\liminf_{n \to \infty} f(x_n) \geq \liminf_{n \to \infty} g_{\bar{p}}(x_n) \geq g_{\bar{p}}(x) = f(x),$$

which completes the proof. □

The algorithm presented by Ahmed et al. [1] is based on the following additional assumptions.

(D4) The technology matrix is fixed, i.e. $T^s = T$ for $s = 1, \dots, S$.

(D5) The first-stage constraint set X is non-empty and compact.

(D6) The recourse matrix W is integral.

Employing assumption (D4), we may reformulate problem (5.3.8) by introducing the variable transformation $\chi = Tx$ for $x \in X$ to obtain the following formulation,

$$\min_{\chi \in \mathcal{X}} \Big\{ F(\chi) = h(\chi) + H(\chi) \Big\} \tag{5.3.9}$$

where

$$h(\chi) = \min\big\{ cx \mid Tx = \chi,\ x \in X \big\},$$
$$H(\chi) = \max_{p \in \mathcal{P}} \Big\{ \overline{\Psi}_p(\chi) = \sum_{s=1}^{S} p_s \Psi^s(\chi) \Big\},$$
$$\Psi^s(\chi) = \min\big\{ q^s y \mid Wy \geq h^s + \chi,\ y \in \mathbb{Z}_+^{n_2} \big\},$$

and

$$\mathcal{X} = \big\{ \chi \in \mathbb{R}^{m_2} \mid \exists\, x \in X : Tx = \chi \big\}.$$

Remark 5.3.8. As in the proof of Proposition 5.3.4 it is easily seen that the function $H(\cdot)$ is real-valued and lower semicontinuous, and since $h(\cdot)$ is clearly real-valued and continuous, we see that problem (5.3.9) is well-defined. Also, observing (D5) we note that the optimal value exists and is actually attained for some $\chi^* \in \mathcal{X}$. Furthermore, given an optimal solution $\chi^* \in \mathcal{X}$ of the transformed problem (5.3.9), it is easily seen that $x^* \in X$ is an optimal solution of the minimax problem (5.3.8) if $x^* \in \arg\min\big\{ cx \mid Tx = \chi^*,\ x \in X \big\}$, cf. also Ahmed et al. [1, Theorem 3.2].

The insight of Ahmed et al. was to note that for any $p \in \mathcal{P}$ the discontinuity points of the function $\overline{\Psi}_p(\cdot)$ are contained in a finite union of hyperplanes that are all orthogonal to the variable axes. In fact, they observe that the function is piecewise constant over rectangular regions of $\mathcal{X}$, the boundaries of which are orthogonal to the variable axes. This result leads directly to the following.

Lemma 5.3.1. *Assume (D1)-(D6), let* $k = (k_1^1, \ldots, k_j^s, \ldots, k_{m_2}^S) \in \mathbb{Z}^{m_2 S}$ *be a vector of integers, and let*

$$C(k) = \bigcap_{s=1}^{S} \prod_{j=1}^{m_2} (k_j^s - h_j^s - 1, k_j^s - h_j^s]$$

and

$$\mathcal{K} = \big\{ k \in \mathbb{Z}^{m_2 S} \mid C(k) \cap \mathcal{X} \neq \emptyset \big\}.$$

Then $|\mathcal{K}| < \infty$ *and for all* $k \in \mathcal{K}$ *the function* $H(\cdot)$ *is constant over* $C(k)$.

Proof. According to [1, Theorem 4.4 and 4.5] the result is true for the function $\overline{\Psi}_p(\cdot)$ for any $p \in \mathcal{P}$. The lemma follows immediately. □

The branch-and-bound algorithm, formally stated below, proceeds by partitioning the feasible set $\mathcal{X}$ into regions of the form $\mathcal{X} \cap \Pi_{j=1}^{m_2}(l_j, u_j]$, where each l_j, $j = 1, \ldots, m_2$, is a possible point of discontinuity of $H(\cdot)$, i.e. $l_j + h_j^s$ is integral for some $s \in \{1, \ldots, S\}$. This is combined with a specialized bounding procedure which is a simple generalization of the one presented by Ahmed et al.

Algorithm 5.3

Step 1 (*Initialization*) Set $\bar{z} = \infty$. Let $l^P, u^P \in \mathbb{R}^{m_2}$ be such that $\mathcal{X} \subseteq \Pi_{j=1}^{m_2} (l_j^P, u_j^P]$ and for all $j = 1, \ldots, m_2$, $l_j^P + h_j^s$ is integral for some $s \in \{1, \ldots, S\}$. Let the list of open problems $\mathcal{L}$ consist of problem P defined by (5.3.9) with the additional constraints $l^P < \chi \leq u^P$. Also, let $\epsilon \in \mathbb{R}_+^{m_2}$ be such that $H(\cdot)$ is constant over $\Pi_{j=1}^{m_2} (l_j, l_j + \epsilon_j]$ whenever $l \in \mathbb{R}^{m_2}$ is such that for all $j = 1, \ldots, m_2$, $l_j + h_j^s$ is integral for some $s \in \{1, \ldots, S\}$.

Step 2 (*Termination/Node selection*) If $\mathcal{L} = \emptyset$, stop; the solution that yielded the upper bound $\bar{z}$ is optimal. Otherwise, select and remove from $\mathcal{L}$ a problem P, defined as $\inf\{F(\chi) \mid l^P < \chi \leq u^P,\ \chi \in \mathcal{X}\}$.

Step 3 (*Bounding*) Obtain a lower bound on P by solving the lower bounding problem $z^P = H(l^P + \epsilon) + \min\{cx \mid Tx = \chi,\ l^P \leq \chi \leq u^P,\ x \in X\}$ and let χ^P be an optimal solution. If $z^P \geq \bar{z}$ go to Step 2. Otherwise, let $\bar{z} = \min\{\bar{z}, F(\chi^P)\}$ and remove from $\mathcal{L}$ all problems P' with $z^{P'} \geq \bar{z}$.

Step 4 (*Branching*) Select an index $j \in \{1, \ldots, m_2\}$ and a value v_j such that $v_j + h_j^s$ is integral for some $s \in \{1, \ldots, S\}$ and $l_j^P < v_j < u_j^P$. Construct two new problems P' and P'', obtained from P by adding the constraints $\chi_j > v_j$ and $\chi_j \leq v_j$, respectively. Let $z^{P'} = z^{P''} = z^P$ and add the two problems to $\mathcal{L}$. Go to Step 2.

Before we prove finite termination of Algorithm 5.3, let us note that Ahmed et al. presented a procedure for the a priori determination of the constant ϵ. The procedure is based on the result in Lemma 5.3.1 and simply determines the smallest possible width of the non-empty regions $C(k)$, $k \in \mathbb{Z}^{m_2 S}$. Observing the definition of ϵ, it is easily seen that the optimal value z^P of the lower bounding problem in Step 3 of the algorithm is a lower bound on the optimal value of the current problem P. In particular, since $H(\cdot)$ is clearly non-decreasing in χ, the definition of ϵ implies that $H(l^P + \epsilon)$ is a lower bound on $H(\chi)$ for $l^P < \chi \leq u^P$. Moreover, $\min\{cx \mid Tx = \chi,\ l^P \leq \chi \leq u^P\ x \in X\}$ is clearly a lower bound on $h(\chi)$ for $l^P < \chi \leq u^P$.

Proposition 5.3.5. *Assume (D1)-(D6). Then Algorithm 5.3 terminates with an optimal solution in a finite number of iterations.*

Proof. Suppose in some iteration of the algorithm that the current problem P is such that $H(\cdot)$ is constant over the set $\{\chi \in \mathcal{X} \mid l^P < \chi \leq u^P\}$. Then $H(\chi^P) \leq H(l^P + \epsilon)$ so that $F(\chi^P) = h(\chi^P) + H(\chi^P) \leq h(\chi^P) + H(l^P + \epsilon) = z^P$ and the current problem is fathomed with no further refinements of the partition. Thus branching only occurs when the set $\{\chi \in \mathcal{X} \mid l^P < \chi \leq u^P\}$ contains a discontinuity point of $H(\cdot)$. By Lemma 5.3.1 and the definition of Step 4, this can only happen a finite number of times and hence the algorithm terminates in a finite number of iterations. Optimality follows from validity of the lower and upper bounding procedures, cf. the proof of [1, Theorem 6.4]. $\square$

For further implementational details, such as e.g. specification of the branching rule and improvements of the lower bounding procedure as well as extension of the algorithm to the case of random technology matrix, we refer to Ahmed, Tawarmalani, and Sahinidis [1].

Part III

Applications in Telecommunications

Chapter 6

Multiperiod Capacity Expansion of One Connection

In this chapter we consider the problem of installing additional capacity on a single telecommunications connection so as to minimize total costs incurred while meeting customer demand. Capacity expansion problems usually give rise to very large and complex models that do not lend themselves to numerical procedures without simplifying assumptions. Capacity expansion of a single telecommunications connection, however, is a relatively simple problem which allows us to include a multiperiod horizon and to abandon the frequently employed assumption of (piecewise) linearity of the cost function, all in all resulting in a relatively exact model. Furthermore, we use a stochastic programming approach when formulating the problem, thus taking due account of the inherent uncertainty involved in the assessment of future demand.

6.1 Two-Stage Formulation

We consider a finite time horizon of T periods, and assume that I different technologies are available for installation in order to supply capacity to meet demand in each period. For $i = 1, \ldots, I$, the capacity supplied by one component of technology i is denoted by c_i and the price of the component is denoted by p_i. The problem is to decide the number of components of each technology to install in each period in order to meet demand. Assuming that demand is known, this problem can be efficiently solved by a procedure proposed by Saniee [136]. In real life, though, the assumption that future demand is known at the point of decision will only in rare cases be justified, and hence demand in each period should rather be thought of as a random variable. It is most natural to think of the probability distribution of this random variable as absolutely continuous, but in practice this would lead to severe computational difficulties. Furthermore, in Chapters 2 and 3 we have seen how stability results in stochastic programming justify the approximation of the true distribution of random parameters by simpler discrete distributions having finite support. Hence we follow this approach, allowing us to think of uncertainty in terms of a number of scenarios, each scenario $s \in \{1, \ldots, S\}$ representing a sequence of outcomes of random demand, $(D_1^s, \ldots, D_T^s)$. Also, we assume that for $s \in \{1, \ldots, S\}$ the probability of scenario s to actually occur is known, and we denote it by π^s.

Not knowing demand at the point of decision means that we may have to accept that we cannot meet all demand in some periods, and hence for $t = 1, \ldots, T$ and $s = 1, \ldots, S$ we denote by z_t^s the amount of demand exceeding capacity in period t under scenario s. Laguna [81] uses this approach to formulate the problem as a robust optimization problem (see Mulvey, Vanderbei, and Zenios [96]), using a general penalty cost function ρ depending on the capacity shortages and the corresponding probabilities of occurrence $\pi^1, \ldots, \pi^S$. If the function ρ is linear, this is essentially a two-stage stochastic programming problem with integer first stage and simple linear recourse. Here the first stage consists of the decisions concerning capacity expansion that must be made without certain knowledge about random demand, whereas the second stage consists of realizations of the capacity shortages that occur once uncertainty is revealed. Denoting by q the cost of one unit of capacity shortage, the two-stage capacity expansion problem is

$$
\begin{aligned}
\min \ & \sum_{t=1}^{T} \gamma^{t-1} \left(\sum_{i=1}^{I} p_i x_{it} + q \sum_{s=1}^{S} \pi^s z_t^s \right) \\
\text{s.t.} \ & \sum_{r=1}^{t} \sum_{i=1}^{I} c_i x_{ir} + z_t^s \geq D_t^s, \quad t = 1, \ldots, T, \ s = 1, \ldots, S, \\
& x_{it} \in \mathbb{Z}_+, \ z_t^s \in \mathbb{R}_+, \quad i = 1, \ldots, I, \ t = 1, \ldots, T, \ s = 1, \ldots, S.
\end{aligned} \tag{6.1.1}
$$

Here γ is a discount factor and for $i = 1, \ldots, I$ and $t = 1, \ldots, T$ we let x_{it} denote the number of components of technology i to be installed in period t. In the following we will assume that both the price and the capacity of every technology are strictly positive.

Remark 6.1.1. The question of actually specifying (or estimating) the cost of lost demand is a complicated matter, since it will only seldom be readily available. If the network operator has the possibility to rent capacity from a competing network operator, q may be taken as the price of such rented capacity. Otherwise q may simply be thought of as lost revenue, possibly with the addition of a penalty cost reflecting customer dissatisfaction. Alternatively, varying values of q may be considered in a parametric analysis illuminating the tradeoff between the cost of the capacity expansion and future capacity shortages.

Remark 6.1.2. Note that if no capacity shortages are allowed, the stochastic programming problem (6.1.1) reduces to a deterministic problem, where for each period $t = 1, \ldots, T$ a demand of $\max_{1 \leq s \leq S} \{D_t^s\}$ is to be satisfied, and hence in this case the problem may be solved by the procedure proposed by Saniee [136].

6.1.1 Solution Procedure

As pointed out in Chapter 2, stochastic programs with simple linear recourse have been the subject of extensive research, and for one thing, efficient solution procedures for such problems have been proposed by various authors. (See e.g. Wets [164].) Clearly, though, the computational difficulties in solving problem (6.1.1) lies not only in its stochastic programming formulation, but to a still larger extent in the integer programming nature of the first-stage problem. To this end, the special structure of the problem was exploited by Laguna [81], who developed a specialized solution procedure, extending the above-mentioned approach of Saniee [136]. Laguna solves problem (6.1.1) in two phases consisting of a sequence of knapsack problems and a shortest path procedure, respectively. In the following we briefly outline this approach.

Let D_{max} denote the maximum demand in any period under any scenario,

$$D_{max} = \max_{1 \le s \le S} \max_{1 \le t \le T} \{D_t^s\}.$$

In the first phase the cost of installing at least y units of capacity in some period is found for $y = 0, \dots, D_{max}$ by solving a series of knapsack problems, defined for a given level y by

$$M(y) = \min\Big\{ \sum_{i=1}^{I} p_i x_i \mid \sum_{i=1}^{I} c_i x_i \ge y,\ x_i \in \mathbb{Z}_+,\ i = 1, \dots, I \Big\}. \tag{6.1.2}$$

Remark 6.1.3. Solving the knapsack problem with right-hand side D_{max} using dynamic programming, produces the desired solutions of $M(y)$ for $y = 1, \dots, D_{max}$. (See e.g. Gilmore and Gomory [49].) In fact Andonov, Poirriez, and Rajopadhye [3] presented a dynamic programming algorithm for the unbounded knapsack problem for which computation time is relatively insensitive to the value of the right-hand side. Hence the series of knapsack problems is solved very efficiently even if some scenario with very large demand exists.

In the second phase Laguna solves a shortest path problem in a directed graph constructed in the following way. The nodes of the graph are ordered in $T + 2$ columns numbered from 0 to $T + 1$. The columns $1, \dots, T$ represent the time periods and each has $D_{max} + 1$ nodes such that the node $v_{t,k}$ represents the situation that at least k units of capacity have been installed by period t. For $t = 2, \dots, T$ and $k = 0, \dots, D_{max}$, all nodes $v_{t-1,k-y}$, $y = 0, \dots, k$, are connected to node $v_{t,k}$ by edges with cost

$$c(v_{t-1,k-y}\,, v_{t,k}) = \gamma^{t-1} M(y) + \gamma^{t-1} q \sum_{s=1}^{S} \pi^s (D_t^s - k)^+, \tag{6.1.3}$$

where $a^+ = \max\{0, a\}$ denotes the positive part of a number $a \in \mathbb{R}$. The first and the last column each has one node, v_0 and v_{T+1}, respectively. v_0 is connected to all nodes $v_{1,k}$, $k = 0, \dots, D_{max}$, by edges with cost

$$c(v_0\,, v_{1,k}) = M(k) + q \sum_{s=1}^{S} \pi^s (D_1^s - k)^+. \tag{6.1.4}$$

Finally, all nodes $v_{T,k}$, $k = 0, \dots, D_{max}$, are connected to v_{T+1} by edges with zero cost. Problem (6.1.1) can now be solved finding the cost of the shortest (v_0, v_{T+1})-path in the graph by some shortest path procedure such as e.g. Dijkstra's algorithm.

Remark 6.1.4. Clearly, the above procedure is easily adapted to account for a more general time-dependency of the penalty cost. In particular, if the penalty cost $\gamma^{t-1} q$ in period t is replaced by q_t for $t = 1, \dots, T$, the edge costs (6.1.3) are simply updated accordingly. If, on the other hand, the unit cost $\gamma^{t-1} p_i$ of technology i in period t is replaced by p_{it} for $i = 1, \dots, I$ and $t = 1, \dots, T$, then a series of knapsack problems (6.1.2) must be solved for each period, and hence for $t = 1, \dots, T$ the term $\gamma^{t-1} M(y)$ in (6.1.3) is replaced accordingly by a more general one, $M_t(y)$.

6.1.2 A New Preprocessing Rule

The computational results reported by Laguna, as well as those presented in Section 6.3 below, seem to indicate that the shortest path procedure is the most time consuming part of the algorithm described above. Because each edge of the graph is considered exactly once during the course of this procedure, an efficient way to reduce computation time is to reduce the number of edges in the graph. With this in mind we make the following observations. First of all note that for $\bar{y} \in \{0, \ldots, D_{max} - 1\}$ the cost $M(\bar{y})$ is associated with the installment of more than $\bar{y}$ units of capacity if and only if $M(\bar{y}) = M(\bar{y}+1)$, and hence in such cases the final term of the edge cost (6.1.3) overestimates the expected penalty cost for lost demand. More specifically we have the following.

Observation 6.1.1. Let $\bar{y} \in \{0, \ldots, D_{max} - 1\}$ be such that $M(\bar{y}) = M(\bar{y}+1)$. Then we have for $t = 2, \ldots, T$ and $k = \bar{y}, \ldots, D_{max} - 1$ that

$$\begin{aligned} c(v_{t-1,k-\bar{y}}\,, v_{t,k}) &= \gamma^{t-1} M(\bar{y}) + \gamma^{t-1} q \sum_{s=1}^{S} \pi^s (D_t^s - k)^+ \\ &\geq \gamma^{t-1} M(\bar{y}+1) + \gamma^{t-1} q \sum_{s=1}^{S} \pi^s (D_t^s - k - 1)^+ \\ &= c(v_{t-1,k-\bar{y}}\,, v_{t,k+1}). \end{aligned}$$

To see how this observation allows us to eliminate from the graph most edges corresponding to the installment of $\bar{y}$ units of capacity, we make the following simple additional observation.

Observation 6.1.2. Because $M(y-1) \leq M(y)$ for any $y \in \{1, \ldots, D_{max}\}$, we have for $t = 1, \ldots, T-1$ and $k, k' \in \{0, \ldots, D_{max}\}$ with $k < k'$ that

$$\begin{aligned} c(v_{t,k}\,, v_{t+1,k'}) &= \gamma^t M(k'-k) + \gamma^t q \sum_{s=1}^{S} \pi^s (D_{t+1}^s - k')^+ \\ &\geq \gamma^t M(k'-k-1) + \gamma^t q \sum_{s=1}^{S} \pi^s (D_{t+1}^s - k')^+ \\ &= c(v_{t,k+1}\,, v_{t+1,k'}). \end{aligned}$$

Let us now consider again the graph that was constructed to solve the capacity expansion problem (6.1.1) as described in the previous section. Employing Observations 6.1.1 and 6.1.2, we see that if $M(\bar{y}) = M(\bar{y}+1)$ for some $\bar{y} \in \{0, \ldots, D_{max} - 1\}$, then the subpath $(v_{t-1,k-\bar{y}}\,, v_{t,k+1}\,, v_{t+1,k'})$ is not longer than the subpath $(v_{t-1,k-\bar{y}}\,, v_{t,k}\,, v_{t+1,k'})$ for $t = 2, \ldots, T-1$, $k = \bar{y}, \ldots, D_{max} - 1$, and $k' = k+1, \ldots, D_{max}$. (The argument does not hold for $k = D_{max}$ because there are no nodes corresponding to the installment of $D_{max}+1$ units of capacity.) Likewise, it is easily seen that for $k' = \bar{y}+1, \ldots, D_{max}$ the subpath $(v_0\,, v_{1,\bar{y}+1}\,, v_{2,k'})$ is not longer than the subpath $(v_0\,, v_{1,\bar{y}}\,, v_{2,k'})$, and for $k = \bar{y}, \ldots, D_{max}-1$ the subpath $(v_{T-1,k-\bar{y}}\,, v_{T,k+1}\,, v_{T+1})$ is not longer than the subpath $(v_{T-1,k-\bar{y}}\,, v_{T,k}\,, v_{T+1})$. Therefore, prior to solving the shortest path problem, we can remove most of the edges corresponding to the installment of $\bar{y}$ units of capacity from the graph, i.e. we remove the edges $(v_0\,, v_{1,\bar{y}})$ and $(v_{t-1,k-\bar{y}}\,, v_{t,k})$ for $t = 2, \ldots, T$ and $k = \bar{y}, \ldots, D_{max} - 1$.

Remark 6.1.5. Note that after application of this preprocessing rule it is easily seen that for $t = 1, \ldots, T$ and $k = 0, \ldots, D_{max} - 1$ the node $v_{t,k}$ now in fact represents the situation that *exactly* k units of capacity have been installed by period t. (For $t = 1, \ldots, T$ the interpretation of the node $v_{t,D_{max}}$, on the other hand, remains unchanged.)

Remark 6.1.6. Clearly, the effect of applying the preprocessing rule is highly dependent on the problem data, and in particular on the prices and capacities of the technologies. Our computational experiments, discussed in Section 6.3, indicates that the reduction in CPU-time closely corresponds to the reduction in the number of edges in the graph.

6.2 Multistage Formulation

It is important at this point to note the distinction between periods and stages. Thus, even though problem (6.1.1) is a multiperiod problem, it only has two stages. As previously pointed out, the first stage includes the decisions on capacity expansion which have to be made without knowing the actual outcome of random demand, whereas the second stage includes realizations of the capacity shortages that occur after uncertainty has been revealed. Because of this mismatch between the number of periods and stages, one might argue that the formulation of the problem given in the previous section does not provide a very good description of the actual process of planning capacity expansion under uncertainty even though the uncertainty of demand is explicitly included. In fact, the formulation in the previous section forces the decision-maker to plan the capacity expansion for the entire time horizon before knowing any outcomes of random demand. It does not seem reasonable, though, that the amount of capacity installed in some period should not depend on actual demand realized up to that period. For this reason one might feel that a multistage formulation of the problem would more appropriately fit the actual multiperiod decision process. For a thorough introduction to the concept of a multistage stochastic recourse program and the related notation, we refer to the textbooks by Birge and Louveaux [22], Kall and Wallace [67], and Prékopa [107], cf. also our discussion in Section 1.2.3.

To facilitate the multistage formulation of the problem, we introduce some additional notation. For $t = 1, \ldots, T$ we denote by $D_{[1,t]}$ a history of demand up to time t,

$$D_{[1,t]} = (D_1, \ldots, D_t), \quad t = 1, \ldots, T.$$

We will refer to $D^s_{[1,t]} = (D^s_1, \ldots, D^s_t)$ as a *subscenario* for $t = 1, \ldots, T$ and $s = 1, \ldots, S$. Also, for each period we define the set of all *distinguishable* subscenarios,

$$\mathcal{S}_t = \left\{ D_{[1,t]} \in \mathbb{R}^t \mid \exists s \in \{1, \ldots, S\} : D_{[1,t]} = D^s_{[1,t]} \right\}, \quad t = 1, \ldots, T,$$

and the corresponding probabilities,

$$\pi(D_{[1,t]}) = \sum_{s : D^s_{[1,t]} = D_{[1,t]}} \pi^s, \quad D_{[1,t]} \in \mathcal{S}_t,\ t = 1, \ldots, T.$$

Finally, we define the set of *descendants* of a subscenario,

$$\mathcal{D}(\bar{D}_{[1,t]}) = \left\{ D_{[1,t+1]} \in \mathcal{S}_{t+1} \mid D_{[1,t]} = \bar{D}_{[1,t]} \right\}, \quad \bar{D}_{[1,t]} \in \mathcal{S}_t,\ t = 1, \ldots, T-1.$$

Remark 6.2.1. For $t = 1, \dots, T$ the subscenarios $D_{[1,t]} \in \mathcal{S}_t$ each corresponds to a bundle of scenarios that are *indistinguishable* at time t, i.e. $\{s \in \{1, \dots, S\} \mid D^s_{[1,t]} = D_{[1,t]}\}$. Hence, the scenarios may be represented by a tree-like structure, with each node representing a subscenario, and with edges connecting subscenarios to their descendants. Such a *scenario tree* was illustrated in Figure 1.1 on page 8.

In the following, we will assume that the decision on capacity expansion for a given period may be based on the development of demand in previous periods. Furthermore, we assume that there is a time delay in the installment of capacity, so that the capacity that we decide to install at some point in time will not be available for use before the beginning of the following period. For $i = 1, \dots, I$, $t = 1, \dots, T$ and $D_{[1,t\text{-}1]} \in \mathcal{S}_{t-1}$, we let $x_{it}(D_{[1,t\text{-}1]})$ denote the number of components of technology i to be installed in period t knowing the demand in periods $1, \dots, t-1$ as realized in subscenario $D_{[1,t\text{-}1]}$. (For ease of notation we define the set $\mathcal{S}_0$ as some arbitrary singleton $\{D_{[1,0]}\}$ and let $\pi(D_{[1,0]}) = 1$ and $\mathcal{D}(D_{[1,0]}) = \mathcal{S}_1$.) Since we have to plan the capacity expansion one period ahead, we may still have to accept that we cannot meet all demand in some periods. For $t = 1, \dots, T$ and $D_{[1,t]} \in \mathcal{S}_t$, we let $z_t(D_{[1,t]})$ denote the amount of demand exceeding capacity in period t under subscenario $D_{[1,t]}$. The multistage capacity expansion problem may now be formulated as

$$
\begin{aligned}
\min \sum_{t=1}^{T} \gamma^{t-1} \Bigg(& \sum_{D_{[1,t\text{-}1]} \in \mathcal{S}_{t\text{-}1}} \pi(D_{[1,t\text{-}1]}) \sum_{i=1}^{I} p_i x_{it}(D_{[1,t\text{-}1]}) + q \sum_{D_{[1,t]} \in \mathcal{S}_t} \pi(D_{[1,t]}) z_t(D_{[1,t]}) \Bigg) \\
\text{s.t.} \sum_{r=1}^{t} \sum_{i=1}^{I} & c_i x_{ir}(D_{[1,r\text{-}1]}) + z_t(D_{[1,t]}) \geq D_t \quad D_{[1,t]} \in \mathcal{S}_t,\ t = 1, \dots, T \\
& x_{it}(D_{[1,t\text{-}1]}) \in \mathbb{Z}_+, \quad D_{[1,t\text{-}1]} \in \mathcal{S}_{t\text{-}1},\ t = 1, \dots, T,\ i = 1, \dots, I. \\
& z_t(D_{[1,t]}) \in \mathbb{R}_+, \quad D_{[1,t]} \in \mathcal{S}_t,\ t = 1, \dots, T.
\end{aligned}
\tag{6.2.1}
$$

Remark 6.2.2. Note that contrary to the two-stage problem (6.1.1), the multistage problem (6.2.1) does not reduce to a deterministic problem when no capacity shortages are allowed. In this case the problem is still a multistage stochastic program, and it can be solved by a recursive solution procedure that is similar in spirit to the one presented for problem (6.2.1) below. We also note that this case corresponds to the situation when there is no time delay in the installment of capacity, so that the capacity installment for some period can be based on the actual demand in that period.

Remark 6.2.3. In the general formulation of a multistage stochastic program presented in Section 1.2.3, we used explicit non-anticipativity constraints, requiring decisions in some period to depend only on the information available at that point. In problem (6.2.1), on the other hand, non-anticipativity is imposed implicitly by allowing decisions to depend only on demand in previous periods. Alternatively, we could define a group of decisions for each scenario and impose explicit non-anticipativity constraints ensuring that

$$
\begin{aligned}
x^s_{it} &= x^{s'}_{it}, \quad i = 1, \dots, I,\ t = 1, \dots, T,\ s, s' \in \{1, \dots, S\},\ D^s_{[1,t-1]} = D^{s'}_{[1,t-1]}, \\
z^s_t &= z^{s'}_t, \quad t = 1, \dots, T,\ s, s' \in \{1, \dots, S\},\ D^s_{[1,t]} = D^{s'}_{[1,t]}.
\end{aligned}
$$

The implicit formulation of non-anticipativity in (6.2.1) is convenient, however, for the recursive solution procedure presented in the following section.

6.2.1 Solution Procedure

Once again we will solve the problem by initially solving the series of knapsack problems (6.1.2) in order to find the cost of installing y units of capacity for $y = 0, \dots, D_{max}$. Now we define the set Y of efficient installment levels cf. Section 6.1.2,

$$Y = \big\{ y \in \{0, \dots, D_{max} - 1\} \mid M(y) < M(y+1) \big\}, \tag{6.2.2}$$

and the related sets

$$Y_k = \big\{ y \in Y \mid 0 \le y \le D_{max} - k \big\} \cup \big\{ D_{max} - k \big\}, \quad k = 0, \dots, D_{max}.$$

Remark 6.2.4. By the same argument that we used in Section 6.1, we will only consider the installment of y units of capacity if $M(y) < M(y+1)$, the only exception being the situation when we choose to install enough capacity to reach a total of D_{max}.

Problem (6.2.1) is now solved by a backward recursion as follows. For $t = T, \dots, 1$, $\bar{D}_{[1,t]} \in \mathcal{S}_t$, and $k = 0, \dots, D_{max}$, we let $Q_t(k, \bar{D}_{[1,t]})$ denote the minimum expected future cost at time t when a total of k units of capacity have already been installed and the history of demand $\bar{D}_{[1,t]}$ has been observed. For the final period we have

$$Q_T(k, \bar{D}_{[1,T]}) = q(\bar{D}_T - k)^+, \quad k = 0, \dots, D_{max},\ \bar{D}_{[1,T]} \in \mathcal{S}_T,$$

and recursively we get

$$\begin{aligned} Q_t(k, \bar{D}_{[1,t]}) = q\big(\bar{D}_t - k\big)^+ & \\ + \gamma \min_{y \in Y_k} \Bigg\{ M(y) + & \sum_{D_{[1,t+1]} \in \mathcal{D}(\bar{D}_{[1,t]})} \frac{\pi(D_{[1,t+1]})}{\pi(\bar{D}_{[1,t]})} Q_{t+1}(k+y, D_{[1,t+1]}) \Bigg\}, \\ & k = 0, \dots, D_{max},\ \bar{D}_{[1,t]} \in \mathcal{S}_t,\ t = T-1, \dots, 1. \end{aligned}$$

The multistage capacity expansion problem can now be stated as

$$\min_{y \in Y_0} \Bigg\{ M(y) + \sum_{D_{[1,1]} \in \mathcal{S}_1} \pi(D_{[1,1]}) Q_1(y, D_{[1,1]}) \Bigg\}.$$

Remark 6.2.5. As noted for the two-stage problem in Section 6.1.1, the recursive solution procedure stated above is easily adapted to account for a more general time-dependency of the penalty cost, whereas a more general time-dependency of the unit cost of each technology requires the solution of a series of knapsack problems (6.1.2) for each period, thus increasing the complexity of the algorithm.

6.3 Computational Experiments

We implemented the algorithm described in Section 6.1 in C++ with and without the new preprocessing rule, primarily to test whether the rule in fact speeds up the algorithm, and secondly to investigate whether the sensitivity of the procedure with respect to I, T, and S is changed when the rule is applied. The experiments were performed on instances of the problem generated randomly according to precepts used by Laguna [81],

$$D_t^s = \sum_{r=1}^{t} d_r^s \qquad t = 1, \dots, T,\ s = 1, \dots, S, \tag{6.3.1a}$$

$$d_t^s \sim N(\mu, \sigma^2) \quad t = 1, \dots, T,\ s = 1, \dots, S, \tag{6.3.1b}$$

$$c_i \sim U(1, \mu) \qquad i = 1, \dots, I, \tag{6.3.1c}$$

$$p_i = \mu + c_i + \tilde{p}_i \quad i = 1, \dots, I, \tag{6.3.1d}$$

$$\tilde{p}_i \sim U(-\sigma, \sigma) \quad i = 1, \dots, I, \tag{6.3.1e}$$

$$\pi^s = \frac{1}{S} \qquad s = 1, \dots, S. \tag{6.3.1f}$$

In compliance with Laguna we generated all instances using $\mu = 100$ and $\sigma = 10$, the discount factor γ was set to 0.86, and the cost of lost demand was fixed at $q = 5$.

The results of the first series of experiments are summarized in Table 6.1. $TSCE1$ refers to the CPU-time (in seconds) used by the procedure without the preprocessing rule and $TSCE2$ refers to the CPU-time when the rule is applied. We also report the maximum demand D_{max} and the number of different solutions from the series of knapsack problems denoted by $|Y|$. (The set Y was defined by (6.2.2).) These numbers are reported because the reason for us to believe that the preprocessing rule is effective, is that it brings the number of installment levels considered in the minimization down from D_{max} to $|Y|$. Finally, we report the reduction in CPU-time and in the number of installment levels considered in the minimization when the new preprocessing rule is applied. All numbers reported are averages over 10 independently generated instances.

Table 6.1: Effect of the Preprocessing Rule

I	10	10	10	10	10	10	4	20	100
T	10	10	10	4	8	12	10	10	10
S	10	100	1000	100	100	100	100	100	100
$TSCE1$	16.23	94.20	892.20	5.74	50.01	161.94	94.22	96.08	95.03
$TSCE2$	2.84	17.08	142.11	0.89	6.53	26.01	9.73	22.57	46.98
Reduction	82.5%	81.9%	84.1%	84.4%	87.0%	83.9%	89.7%	76.5%	50.6%
D_{max}	1034.8	1081.4	1106.0	446.6	867.8	1287.3	1082.8	1084.7	1078.3
$\|Y\|$	174.6	215.5	187.6	72.2	124.2	218.2	113.0	270.2	552.0
Reduction	83.1%	80.1%	83.0%	83.8%	85.7%	83.0%	89.6%	75.1%	48.8%

From Table 6.1 we see that the algorithm provides consistently shorter CPU-times when the new preprocessing rule is applied, the reduction in CPU-time closely corresponding to the reduction in the number of installment levels considered in the minimization. Also, as reported by Laguna, our experiments show that the procedure is quite sensitive with respect to the number of periods T as well as the number of scenarios S, whereas the algorithm seems to be rather insensitive with respect to the number of technologies I when the preprocessing rule is not applied, all in all indicating that the CPU-time needed to solve the series of knapsack problems is negligible compared to the CPU-time needed for the shortest path procedure. When the preprocessing rule is applied, however, we note an increase in CPU-time for increasing numbers of technologies. This can be ascribed to the fact that an increasing number of technologies implies an increasing number of different solutions from the series of knapsack problems as indicated by $|Y|$.

The recursive solution procedure for the multistage problem described in Section 6.2 was also implemented in C++ and a series of experiments were carried out. The purpose of these experiments was twofold. First, we wanted to test if the multistage method is practicable. Second, the experiments provide some insight into the difference between the two-stage and the multistage models. To generate scenario trees for the multistage problem we considered a fixed number of $\bar{S}$ descendants for all subscenarios, resulting in a total of $S = \bar{S}^T$ scenarios. Again the random instances were generated using the precepts (6.3.1). When generating demand by (6.3.1a) and (6.3.1b), though, we now had to collect the scenarios in scenario bundles in each period. This was done by generating $\bar{S}$ demand values for period 1, for each of these values $\bar{S}$ independent demand increments (6.3.1b) were generated and so forth.

Results of the second series of computational experiments are reported in Table 6.2. Again we report average results from 10 independently generated instances. The CPU-time used by the solution procedure for the multistage problem is referred to as $MSCE$. For each instance we also ran a modified solution procedure for the two-stage problem, taking advantage of the bundling of scenarios in the scenario tree, i.e. simply replacing the term $\sum_{s=1}^{S} \pi^s (D_t^s - k)^+$ in the edge cost (6.1.3) or (6.1.4) by $\sum_{D_{[1,t]} \in \mathcal{S}_t} \pi(D_{[1,t]})(D_t - k)^+$ for $t = 1, \ldots, T$. We refer to the CPU-time used by this procedure as $TSCE3$.

Table 6.2: Practicability of the Multistage Algorithm

I	10	10	10	10	10	10	4	20	100
T	6	6	6	4	8	12	10	10	10
$\bar{S}$	2	3	4	2	2	2	2	2	2
$MSCE$	1.82	16.81	76.06	0.20	14.35	687.08	36.71	150.50	309.32
$TSCE3$	0.84	6.28	29.84	0.14	5.20	215.47	11.80	48.54	100.68

From Table 6.2 we see that the multistage model is certainly practicable even though the solution time is up to about three times larger than that of the two-stage model. The increased solution times are a consequence of the improved accuracy provided by the multistage model. We note, however, that the cost savings achieved by using a multistage model were rather small for the instances considered here (less than 3%), and in fact the two-stage model and the multistage model provided identical first-period decisions for several instances. Hence if time is a critical factor, it may be that one can settle for the results of the two-stage model with no grave consequences. Clearly, though, this conjecture cannot be confirmed without an extensive amount of computational testing concerning the particular instance at hand.

Acknowledgment

The author would like to thank David Pisinger for providing the software for solution of knapsack problems.

Chapter 7

Capacitated Network Design

In this chapter we consider the design of a capacitated telecommunications network. The problem is to install additional capacity on the transmission links and subsequently route traffic in the resulting network so as to meet customer demand while minimizing total costs incurred. We will assume that two facilities with fixed capacities are available for installation but most of the results may be generalized in case several facilities are available. The polyhedral structure of the two-facility capacitated network design problem with known point-to-point traffic demands has been studied by e.g. Bienstock and Günlük [17] and Günlük [53]. Here the authors derive facet-defining inequalities and use these to solve the problem by cutting plane procedures. It is a trivial observation, though, that uncertainty is almost always an inherent feature of systems involving the assessment of future demand, and this is particularly so in the present context, since telecommunications is a branch of trade that is currently undergoing a thriving development with rapidly increasing demand. In this chapter, following a brief survey of well-known results for the deterministic capacitated network design problem, we consider the alternative of formulating the problem as a two-stage stochastic program with integer first stage and continuous second stage, hence taking due account of the inherent uncertainty involved in the assessment of future demand.

7.1 The Deterministic Problem

The network is modeled as a connected undirected graph $G = (V, E)$ with the node set V representing switches in the network and the edge set E representing transmission links. In this network a number of directed point-to-point traffic demands are to be routed. To this end, demand will be described by a set K of commodities, and for $k \in K$ and $i \in V$ we let D_{ik} denote the net demand for commodity k at node i.

Remark 7.1.1. Several possibilities exist for defining the demand commodities. A common approach is to define a commodity for every non-zero point-to-point demand, resulting in a total of $O(|V|^2)$ commodities. To reduce the number of variables and constraints, however, we will prefer to work here with an aggregated formulation in which each commodity corresponds to the set of all point-to-point demands originating at some particular node, hence resulting in a total of only $O(|V|)$ commodities. In other words, we assume throughout the following that $K \subseteq V$.

The existing capacity on an edge $\{i,j\} \in E$ is denoted by C_{ij}. Since we are modeling a communications network with optical transmission systems we will assume that a given edge in the network can carry flow in either direction and, more importantly, that these flows do not interfere. Hence, each undirected edge $\{i,j\} \in E$ conceptually corresponds to two directed edges (i,j) and (j,i), both with capacity C_{ij}. Additional capacity may be installed on edges of the network in multiples of two batch sizes corresponding to low-capacity and high-capacity facilities, respectively. We will assume that the smaller batch size is equal to 1, which may be achieved by rescaling demand. The larger batch size will be denoted by λ and we will assume that λ is an integer. We are now ready to define the convex hull of feasible solutions for the capacitated network design problem,

$$\mathcal{P} = \operatorname{conv}\Bigg\{ (x,y,f) \in \mathbb{Z}_+^{|E|} \times \mathbb{Z}_+^{|E|} \times \mathbb{R}_+^{2|K||E|} \;\Bigg|$$

$$\sum_{j:\{i,j\}\in E} f_{jik} - \sum_{j:\{i,j\}\in E} f_{ijk} = D_{ik}, \qquad i \in V,\ k \in K,\ i \neq k, \tag{7.1.1a}$$

$$\sum_{k\in K} f_{ijk} \leq C_{ij} + x_{ij} + \lambda y_{ij}, \qquad \{i,j\} \in E, \tag{7.1.1b}$$

$$\sum_{k\in K} f_{jik} \leq C_{ij} + x_{ij} + \lambda y_{ij}, \qquad \{i,j\} \in E \Bigg\}. \tag{7.1.1c}$$

Here, for $\{i,j\} \in E$ we denote by x_{ij} (y_{ij}) the number of low-capacity (high-capacity) facilities to be installed on edge $\{i,j\}$, and for $k \in K$ we denote by f_{ijk} and f_{jik} the flow of commodity k on the two conceptual edges (i,j) and (j,i) corresponding to edge $\{i,j\}$. Equation (7.1.1a) is a flow conservation constraint while (7.1.1b) and (7.1.1c) are capacity constraints. In the following, whenever $e = \{i,j\} \in E$ we will refer to the same variable by x_{ij}, x_{ji}, and x_e interchangeably. A similar notation will be used for existing capacity and for high-capacity facilities.

For $\{i,j\} \in E$ the cost of installing a low-capacity facility on the edge $\{i,j\}$ is denoted by a_{ij} and the corresponding cost of a high-capacity facility is denoted by b_{ij}. Finally, for $\{i,j\} \in E$ and $k \in K$ we will use a flow cost $c_{ijk} = c_{jik}$ for one unit of commodity k on the edge $\{i,j\}$, knowing that this cost may well be zero in many real-life applications. The deterministic capacitated network design problem may now be stated as

$$\min\big\{ax + by + cf \mid (x,y,f) \in \mathcal{P}\big\}, \tag{7.1.2}$$

where a, b, and c are the cost vectors, and transposes have once again been omitted for simplicity.

Remark 7.1.2. The closely related network loading problem has been considered by Magnanti, Mirchandani, and Vachani [88, 89], and Mirchandani [93], who obtained results similar to those presented for the capacitated network design problem in the following sections. The network loading problem is a slight variation of the capacitated network design problem in which it is assumed that there is no existing capacity on the edges of the network and no cost of flow. We note that the capacitated network design problem will generally be relevant for the network provider, whereas the network loading problem typically arises when customers wish to design a private-line network by leasing transmission facilities from the network provider.

7.1.1 Metric Inequalities

The class of so-called metric inequalities, originally introduced by Iri [63] and Onaga and Kakusho [104], can be used to project the feasible region $\mathcal{P}$ onto the space of discrete capacity variables. These valid inequalities have been applied in the context of capacitated network design by various authors. Günlük [53] briefly describes the inequalities, but argues that it is only practical to use special subclasses of them in cutting plane procedures for problem (7.1.2). Bienstock et al. [16], on the other hand, employs metric inequalities in a cutting plane procedure for a reformulation of the one-facility capacitated network design problem using flow variables related to a number of prespecified paths rather than the edge-related flow variables used in the formulation (7.1.1). For the stochastic programming problem that we are going to consider, the metric inequalities will be used as feasibility cuts in an L-shaped algorithm and hence we consider them in some detail.

For fixed values of x and y, problem (7.1.2) is a standard multicommodity flow problem. Associating dual variables ρ, σ, and τ with constraints (7.1.1a), (7.1.1b), and (7.1.1c), respectively, the dual of this problem is

$$\begin{aligned} \max \ & \sum_{i\in V}\sum_{k\in K} D_{ik}\rho_{ik} - \sum_{\{i,j\}\in E} (\sigma_{ij}+\tau_{ij})(C_{ij}+x_{ij}+\lambda y_{ij}) \\ \text{s.t. } & \rho_{jk}-\rho_{ik}-\sigma_{ij} \le c_{ijk}, \quad \{i,j\}\in E,\ k\in K, \\ & \rho_{ik}-\rho_{jk}-\tau_{ij} \le c_{ijk}, \quad \{i,j\}\in E,\ k\in K, \\ & \rho_{kk}=0, \quad k\in K, \\ & \rho\in\mathbb{R}^{|V||K|},\ \sigma,\tau\in\mathbb{R}_+^{|E|}. \end{aligned} \tag{7.1.3}$$

The multicommodity flow problem is feasible if and only if the dual problem (7.1.3) is bounded. That is, if and only if the following inequalities hold,

$$\sum_{\{i,j\}\in E} (v_{ij}+w_{ij})(C_{ij}+x_{ij}+\lambda y_{ij}) \ge \sum_{i\in V}\sum_{k\in K} D_{ik}u_{ik}, \quad (u,v,w)\in\mathcal{D}^+, \tag{7.1.4}$$

where $\mathcal{D}^+$ denotes the recession cone of the feasible region of the dual problem,

$$\begin{aligned} \mathcal{D}^+ = \Big\{ & (u,v,w)\in\mathbb{R}^{|V||K|}\times\mathbb{R}_+^{|E|}\times\mathbb{R}_+^{|E|} \ \Big| \\ & u_{jk}-u_{ik}\le v_{ij},\ u_{ik}-u_{jk}\le w_{ij},\ u_{kk}=0,\ \{i,j\}\in E,\ k\in K \Big\}. \end{aligned} \tag{7.1.5}$$

The inequalities defined by (7.1.4) are referred to as *metric inequalities*.

Remark 7.1.3. To understand why the inequalities defined by (7.1.4) are called metric inequalities, consider a directed graph $\bar{G}=(V,A)$ where the edge set A contains the two directed edges (i,j) and (j,i) for each undirected edge $\{i,j\}\in E$. Now, associating weights v_{ij} with edge (i,j) and w_{ij} with edge (j,i) for each $\{i,j\}\in E$, it is easily seen that the right-hand side of (7.1.4) is maximized if and only if u_{ik} is the length of a shortest (k,i)-path in $\bar{G}$ for all $k\in K\subseteq V$ and $i\in V$. From now on we will only be interested in the extreme rays $(u,v,w)\in\mathcal{D}^+$ of the feasible region in (7.1.3), that satisfies this property.

Since the feasible region of the dual problem (7.1.3) is a rational polyhedron, we may assume that the extreme rays $(u, v, w) \in \mathcal{D}^+$ are integral — this can be achieved by scaling. Hence the metric inequalities (7.1.4) may be strengthened by rounding,

$$\sum_{\{i,j\}\in E} (v_{ij} + w_{ij})(x_{ij} + \lambda y_{ij}) \geq \left\lceil \sum_{i\in V}\sum_{k\in K} D_{ik}u_{ik} - \sum_{\{i,j\}\in E} C_{ij}(v_{ij} + w_{ij}) \right\rceil, \ (u, v, w) \in \mathcal{D}^+.$$

The strengthened inequalities are referred to as *integral metric inequalities*. They do not necessarily define facets of $\mathcal{P}$ but, as already mentioned, they will be useful as feasibility cuts in an L-shaped algorithm for a stochastic programming formulation of the problem.

7.1.2 Partition Inequalities

In this section we consider a special subclass of integral metric inequalities referred to as *partition inequalities*. The reason for us to consider these inequalities separately is that some partition inequalities are facet-defining under mild conditions. Here we will only consider partition inequalities obtained as integral metric inequalities by assigning unit weight to some edges and zero weight to the remaining edges cf. Remark 7.1.3.

Let $\pi = (V_1, \ldots, V_l)$ be a partition of the node set into l subsets and let E_π denote the corresponding multicut, $E_\pi = \big\{\{i,j\} \in E \,\big|\, \exists\, r \in \{1, \ldots, l\} \text{ such that } |\{i,j\} \cap V_r| = 1\big\}$. Next, for any permutation $\alpha = (\alpha_1, \ldots, \alpha_l)$ of the sequence $(1, \ldots, l)$, we let $T(\pi, \alpha)$ denote the net traffic that must be routed across the multicut E_π from lower-numbered subsets to higher-numbered subsets when subsets are numbered with respect to α,

$$T(\pi, \alpha) = \sum_{r=1}^{l-1} \sum_{k\in V_{\alpha_r}} \sum_{i\in V^{\alpha_r}} D_{ik} - \sum_{e\in E_\pi} C_e,$$

where $V^{\alpha_r} = \bigcup_{t=r+1}^{l} V_{\alpha_t}$. $T(\pi, \alpha)$ provides a lower bound on the capacity that must be installed across the multicut E_π. Taking the maximum over the $l!$ possible permutations of $(1, \ldots, l)$, we obtain a stronger lower bound $T(\pi)$ and a valid inequality for $\mathcal{P}$,

$$x(E_\pi) + \lambda y(E_\pi) \geq \lceil T(\pi) \rceil, \tag{7.1.6}$$

where we define $x(E_\pi) = \sum_{e\in E_\pi} x_e$ and $y(E_\pi) = \sum_{e\in E_\pi} y_e$ for notational convenience.

Remark 7.1.4. Note that the valid inequality with left-hand side as in (7.1.6) and right-hand side equal to $\lceil T(\pi, \alpha) \rceil$ is an integral metric inequality obtained in the following way. For each undirected edge $\{i,j\} \in E_\pi$ let $A_\pi \subseteq A$ contain the directed edge (i,j) or (j,i) going from a lower-numbered subset to a higher-numbered subset when subsets are numbered with respect to α. The desired inequality is now obtained by assigning unit weights to edges in A_π and zero weight to all other edges in A cf. Remark 7.1.3.

Remark 7.1.5. The inequality defined by (7.1.6) is referred to as an *l-partition inequality*. 2-partition inequalities are usually referred to as *cutset inequalities*. A cutset inequality defines a facet of $\mathcal{P}$ provided that V_1 as well as $V_2 = V \setminus V_1$ are not empty and induce connected subgraphs, and that $T(V_1, V_2)$ is not integer. A proof of this result may be found in e.g. Bienstock and Günlük [17], who also give several sufficient conditions for

3-partition inequalities to be facet-defining. Finally, Bienstock et al. [16] consider a one-facility capacitated network design problem and give sufficient conditions for a general l-partition inequality to be facet-defining for the projection of the feasible region on the space of discrete capacity variables.

7.1.3 Mixed-Integer Rounding Inequalities

Applying the mixed-integer rounding procedure (see e.g. Nemhauser and Wolsey [97]) to the partition inequalities, we obtain a new class of valid inequalities referred to as *mixed-integer rounding inequalities*. For notational convenience we let $\bar{T}_\pi$ denote the right-hand side of the partition inequality (7.1.6) for a given partition π, that is $\bar{T}_\pi = \lceil T(\pi) \rceil$, and we let $r_\pi = (\bar{T}_\pi \bmod \lambda)$. Note that $\bar{T}_\pi = \lambda \lfloor \bar{T}_\pi/\lambda \rfloor + r_\pi$ with $0 \leq r_\pi < \lambda$, and $r_\pi = 0$ if and only if $\bar{T}_\pi$ is a scalar multiple of λ. In terms of the aggregate variables $x(E_\pi)$ and $y(E_\pi)$ the partition inequality (7.1.6) is tight at integer points $(\bar{T}_\pi, 0), (\bar{T}_\pi - \lambda, 1), \ldots, (r_\pi, \lfloor \bar{T}_\pi/\lambda \rfloor)$. Still, when $\bar{T}_\pi$ is not a scalar multiple of λ, new fractional extreme points with $x(E_\pi) = 0$ and $y(E_\pi) = \bar{T}_\pi/\lambda$ are induced. Such points may be cut off by the following mixed-integer rounding inequality,

$$x(E_\pi) + r_\pi y(E_\pi) \geq r_\pi \lceil \bar{T}_\pi/\lambda \rceil. \tag{7.1.7}$$

In terms of the aggregate variables $x(E_\pi)$ and $y(E_\pi)$ this inequality is tight at integer points $(0, \lceil \bar{T}_\pi/\lambda \rceil)$ and $(r_\pi, \lfloor \bar{T}_\pi/\lambda \rfloor)$. Thus, whenever $r_\pi \neq 0$, the mixed-integer rounding inequality (7.1.7) is stronger than the corresponding partition inequality for points with $\lfloor \bar{T}_\pi/\lambda \rfloor \leq y(E_\pi) \leq \lceil \bar{T}_\pi/\lambda \rceil$. If $r_\pi = 0$, on the other hand, the inequality is redundant.

Remark 7.1.6. It is possible to prove that if a partition inequality (7.1.6) is facet-defining for $\mathcal{P}$, then the corresponding mixed-integer rounding inequality (7.1.7) is also facet-defining for $\mathcal{P}$ under mild conditions cf. e.g. Bienstock and Günlük [17].

7.1.4 Mixed Partition Inequalities

Finally, we consider the class of *mixed partition inequalities*, introduced by Günlük [53]. Let π and π' be two distinct partitions of the node set V and consider the related partition inequalities (7.1.6). Again we let $\bar{T}_\pi$ and $\bar{T}_{\pi'}$ denote the right-hand sides, and r_π and $r_{\pi'}$ the corresponding remainders by division with λ. Assuming that $r_\pi > r_{\pi'}$, and applying a general mixing procedure for mixed-integer sets, introduced by Günlük and Pochet [54], to the two partition inequalities, we obtain a valid inequality for $\mathcal{P}$,

$$x(E_\pi \cup E_{\pi'}) + (r_\pi - r_{\pi'}) y(E_\pi) + r_{\pi'} y(E_{\pi'}) \geq (r_\pi - r_{\pi'}) \lceil \bar{T}_\pi/\lambda \rceil + r_{\pi'} \lceil \bar{T}_{\pi'}/\lambda \rceil.$$

Remark 7.1.7. Günlük [53] presented several conditions for mixed partition inequalities to be facet-defining for $\mathcal{P}$, but his computational experiments indicated that the inequalities only in rare cases had an effect when partition inequalities and mixed-integer rounding inequalities were already included in the formulation. It turns out, however, that mixed partition inequalities are the only class of inequalities presented so far that may combine information from different scenarios in the stochastic programming formulation, and hence they turned out to be quite useful.

7.2 The Stochastic Programming Problem

Until now we have only considered the deterministic capacitated network design problem. It is worth noting at this point, though, that the number of low- and high-capacity facilities to install on each edge of the network must be decided upon well in advance of the point in time at which they are actually installed and operating. Thus, the assumption that demand is known at the point of decision will generally not be justified since uncertainty is almost always an inherent feature of systems involving the assessment of future demand. Furthermore, it is quite likely that the capacity expansion is required to be sufficient for some period of time causing even more uncertainty about the actual volume of demand that needs to be satisfied. Therefore we now propose an alternative stochastic programming formulation of the capacitated network design problem. In particular, we formulate the problem as a two-stage stochastic program with integer first stage and continuous second stage. Here the first stage involves the installment of capacity that must be planned here-and-now, based solely on the information about demand conveyed through its distribution, whereas the second stage involves the routing of traffic which is naturally postponed until the actual outcome of random demand is realized.

Remark 7.2.1. Before we proceed, it may be appropriate to point out in more detail how uncertainty may arise in our model. Usually the arrival of demands (transmissions) is described by a Poisson process, the length of a transmission is described by an exponentially distributed random variable, and even the required bandwidth of a transmission may be a random variable. Also, it is not (economically) feasible to construct a bufferless network in which no blocking will occur even in extreme peak situations. Dempster, Medova, and Thompson [40], and Medova [90] introduced a possible approach in the context of ATM-based broadband integrated services digital networks (B-ISDN). Here the capacity expansion is carried out subject to certain grade-of-service (GoS) constraints, corresponding to a certain set of blocking probabilities that must not be exceeded. Given the blocking probabilities and the distributions that describe demand, it is possible to determine the so-called *effective bandwidth* requirements, which can be thought of as the capacity needed to ensure that the blocking probabilities are not exceeded. (For more details on the actual computation of effective bandwidth requirements we refer to Medova [90].) The effective bandwidth requirements serve as demand input for our problem and a feasible solution is required to observe these requirements so that the blocking probabilities are not exceeded and the GoS is maintained. Uncertainty in our formulation of the problem arises due to the fact that the distributions describing future demand are generally unknown and may at best be replaced by approximations based for example on historical data and some sort of forecast model.

As previously pointed out also in Chapter 6, it is most natural to think of the probability distribution of future demand as absolutely continuous, but in practice this would lead to severe computational difficulties. Hence we follow a scenario approach, representing the uncertain outcome of future demand by a finite number of scenarios. Associated with each scenario $s \in \{1, \dots, S\}$ is a realization of random demand D^s, a corresponding routing of traffic f^s, and a cost of flow c^s. Finally, we assume that for $s \in \{1, \dots, S\}$ the probability of scenario s to actually occur is known, and we denote it by p^s.

Remark 7.2.2. In fact, as described in Section 7.4.2, scenarios were generated, not in an attempt to approximate some true probability distribution of random parameters, but merely to represent the spectrum of possibly future outcomes of random demand. Hence, some may say that our stochastic programming model should rather be referred to as a *robust optimization* problem (see Mulvey, Vanderbei, and Zenios [96]), even though we require the second-stage problem to be feasible for all scenarios and hence do not consider the use of a penalty cost function for second-stage infeasibilities.

The feasible region of the stochastic capacitated network design problem, when only flow under some scenario $s \in \{1, \ldots, S\}$ is restricted, is

$$\mathcal{R}^s = \Bigg\{ (x, y, f^1, \ldots, f^S) \in \mathbb{Z}_+^{|E|} \times \mathbb{Z}_+^{|E|} \times \mathbb{R}_+^{2|K||E|} \times \cdots \times \mathbb{R}_+^{2|K||E|} \;\Bigg|$$

$$\sum_{j:\{i,j\}\in E} f_{jik}^s - \sum_{j:\{i,j\}\in E} f_{ijk}^s = D_{ik}^s, \qquad i \in V,\ k \in K,\ i \neq k, \tag{7.2.1a}$$

$$\sum_{k\in K} f_{ijk}^s \leq C_{ij} + x_{ij} + \lambda y_{ij}, \qquad \{i,j\} \in E, \tag{7.2.1b}$$

$$\sum_{k\in K} f_{jik}^s \leq C_{ij} + x_{ij} + \lambda y_{ij}, \qquad \{i,j\} \in E \Bigg\}, \tag{7.2.1c}$$

and the feasible region of the stochastic programming problem is

$$\mathcal{R} = \bigcap_{s=1}^{S} \mathcal{R}^s.$$

Now we may state the capacitated network design problem as a two-stage stochastic programming problem in which the sum of total installment cost and expected flow cost is minimized subject to the usual flow conservation and capacity constraints,

$$\min\Bigg\{ ax + by + \sum_{s=1}^{S} p^s c^s f^s \;\Bigg|\; (x, y, f^1, \ldots, f^S) \in \operatorname{conv} \mathcal{R} \Bigg\}. \tag{7.2.2}$$

For the following analysis, it will be convenient to reformulate (7.2.2) in terms of the capacity variables x and y only. To this end we define the projections of the sets $\mathcal{R}$ and $\mathcal{R}^s$, $s = 1, \ldots, S$, on the space of discrete capacity variables x and y,

$$\mathcal{R}_{x,y}^s = \Big\{ (x, y) \in \mathbb{Z}_+^{|E|} \times \mathbb{Z}_+^{|E|} \;\Big|\; \exists f^1, \ldots, f^S \in \mathbb{R}_+^{2|K||E|} \,:\, (x, y, f^1, \ldots, f^S) \in \mathcal{R}^s \Big\},$$

$$\mathcal{R}_{x,y} = \bigcap_{s=1}^{S} \mathcal{R}_{x,y}^s.$$

Problem (7.2.2) may now be equivalently stated as

$$\min\Bigg\{ ax + by + \sum_{s=1}^{S} p^s Q^s(x, y) \;\Bigg|\; (x, y) \in \operatorname{conv} \mathcal{R}_{x,y} \Bigg\}, \tag{7.2.3}$$

where the second-stage value function $Q^s(x, y)$ is defined for each $s \in \{1, \ldots, S\}$ by

$$Q^s(x, y) = \min\big\{ c^s f^s \mid (x, y, f^1, \ldots, f^S) \in \mathcal{R}^s \big\}. \tag{7.2.4}$$

7.2.1 Valid Inequalities

Comparing the structure of the regions $\mathcal{P}$ and $\operatorname{conv}\mathcal{R}^s$ for some $s \in \{1,\dots,S\}$, it is obvious that we may obtain integral metric inequalities, partition inequalities, and mixed-integer rounding inequalities for $\operatorname{conv}\mathcal{R}^s$ in the exact same way as discussed for the deterministic problem in previous sections.

For $s \in \{1,\dots,S\}$, consider the multicommodity flow problem obtained by fixing the values of the capacity variables x and y and minimizing the flow costs subject to (7.2.1a), (7.2.1b), and (7.2.1c). Note that the recession cone $\mathcal{D}^+$ defined by (7.1.5) is the same for all of these problems. Hence for any dual extreme ray $(u,v,w) \in \mathcal{D}^+$ we obtain for each $s \in \{1,\dots,S\}$ an integral metric inequality for $\operatorname{conv}\mathcal{R}^s$,

$$\sum_{\{i,j\}\in E} (v_{ij}+w_{ij})(x_{ij}+\lambda y_{ij}) \geq \left\lceil \sum_{i\in V}\sum_{k\in K} D^s_{ik}u_{ik} - \sum_{\{i,j\}\in E} C_{ij}(v_{ij}+w_{ij}) \right\rceil. \qquad (7.2.5)$$

Next, for any partition $\pi = (V_1,\dots,V_l)$ of the node set we may calculate the maximum net traffic that needs to be routed across the multicut E_π under each scenario $s \in \{1,\dots,S\}$,

$$T^s(\pi) = \max_{\alpha=(\alpha_1,\dots,\alpha_l)} \left\{ \sum_{r=1}^{l-1} \sum_{k\in V_{\alpha_r}} \sum_{i\in V^{\alpha_r}} D^s_{ik} \right\} - \sum_{e\in E_\pi} C_e,$$

cf. the discussion in Section 7.1.2. Hence for $s \in \{1,\dots,S\}$ we let $\bar{T}^s_\pi = \lceil T^s(\pi) \rceil$ to obtain an l-partition inequality for $\operatorname{conv}\mathcal{R}^s$,

$$x(E_\pi) + \lambda y(E_\pi) \geq \bar{T}^s_\pi, \qquad (7.2.6)$$

and we let $r^s_\pi = (\bar{T}^s_\pi \bmod \lambda)$ to obtain a mixed-integer rounding inequality for $\operatorname{conv}\mathcal{R}^s$,

$$x(E_\pi) + r^s_\pi y(E_\pi) \geq r^s_\pi \lceil \bar{T}^s_\pi / \lambda \rceil. \qquad (7.2.7)$$

As pointed out, these three classes of inequalities are valid for $\operatorname{conv}\mathcal{R}^s$, $s \in \{1,\dots,S\}$, and hence for $\operatorname{conv}\mathcal{R}$. Thus in principle we may generate cuts of each type from all of the S scenarios. It is easily seen, though, that the similarities of cuts generated from different scenarios should be exploited. In particular, it is only natural to consider the partition inequality (7.2.6) with maximum right-hand side, and the corresponding mixed-integer rounding inequality (7.2.7). Hence we define

$$\bar{T}_\pi = \bar{T}^{s^*}_\pi \quad \text{where} \quad s^* \in \arg\max_{1\leq s\leq S}\{\bar{T}^s_\pi\},$$
$$r_\pi = \bar{T}_\pi \bmod \lambda.$$

Proposition 7.2.1. *For any scenario $s \in \{1,\dots,S\}$ the partition inequality (7.2.6) is dominated by the partition inequality*

$$x(E_\pi) + \lambda y(E_\pi) \geq \bar{T}_\pi. \qquad (7.2.8)$$

Proof. The result is obvious. □

Obviously, a similar result may be stated for the integral metric inequalities (7.2.5). When we turn to mixed-integer rounding inequalities, on the other hand, a bit more care must be taken, since these cuts are not parallel for different scenarios.

Proposition 7.2.2. *For any scenario $s \in \{1, \dots, S\}$ the mixed-integer rounding inequality (7.2.7) is dominated by either non-negativity constraints, the partition inequality (7.2.8), or the mixed-integer rounding inequality*

$$x(E_\pi) + r_\pi y(E_\pi) \geq r_\pi \lceil \bar{T}_\pi / \lambda \rceil. \tag{7.2.9}$$

Proof. Let $s \in \{1, \dots, S\}$ and write the corresponding mixed-integer rounding inequality (7.2.7) as

$$x(E_\pi) \geq r_\pi^s \big(\lceil \bar{T}_\pi^s / \lambda \rceil - y(E_\pi) \big). \tag{7.2.10}$$

Similarly, write (7.2.8) and (7.2.9) as

$$x(E_\pi) \geq \bar{T}_\pi - \lambda y(E_\pi), \tag{7.2.11}$$

and

$$x(E_\pi) \geq r_\pi \big(\lceil \bar{T}_\pi / \lambda \rceil - y(E_\pi) \big), \tag{7.2.12}$$

respectively. First of all we note that unless $0 \leq y(E_\pi) \leq \lceil \bar{T}_\pi^s / \lambda \rceil$, the inequality (7.2.10) is dominated by non-negativity constraints. Next, we note that if $r_\pi^s \leq r_\pi$, the inequality (7.2.10) is dominated by (7.2.12) whenever $0 \leq y(E_\pi) \leq \lceil \bar{T}_\pi^s / \lambda \rceil$. So assume that $r_\pi^s > r_\pi$. Now, since $\lceil \bar{T}_\pi^s / \lambda \rceil = \lceil \bar{T}_\pi / \lambda \rceil$ implies $r_\pi^s \leq r_\pi$, we must have $\lceil \bar{T}_\pi^s / \lambda \rceil \leq \lfloor \bar{T}_\pi / \lambda \rfloor$, and hence (7.2.10) is dominated by

$$x(E_\pi) \geq r_\pi^s \big(\lfloor \bar{T}_\pi / \lambda \rfloor - y(E_\pi) \big). \tag{7.2.13}$$

To see that this inequality is dominated by the partition inequality (7.2.11), we use the fact that $\bar{T}_\pi = r_\pi + \lambda(\lfloor \bar{T}_\pi / \lambda \rfloor)$ to write (7.2.11) as

$$x(E_\pi) \geq r_\pi + \lambda \big(\lfloor \bar{T}_\pi / \lambda \rfloor - y(E_\pi) \big).$$

Since $r_\pi^s < \lambda$ and $r_\pi \geq 0$, we see that for $0 \leq y(E_\pi) \leq \lceil \bar{T}_\pi^s / \lambda \rceil$ the inequality (7.2.11) dominates (7.2.13) and with that also (7.2.10). $\square$

Finally we turn to the class of mixed partition inequalities. Unlike the previously considered inequalities, we may derive inequalities of this type combining information from different scenarios. Thus, let us consider two maximal partition inequalities (7.2.8) corresponding to two distinct partitions π and π'. We denote by $\bar{T}_\pi$, $\bar{T}_{\pi'}$, r_π, and $r_{\pi'}$ the right-hand sides and corresponding remainders by division with λ, and we assume once again that $r_\pi > r_{\pi'}$. Applying the mixing procedure of Günlük and Pochet [54] to these inequalities we obtain a mixed partition inequality,

$$x(E_\pi \cup E_{\pi'}) + (r_\pi - r_{\pi'}) y(E_\pi) + r_{\pi'} y(E_{\pi'}) \geq (r_\pi - r_{\pi'}) \lceil \bar{T}_\pi / \lambda \rceil + r_{\pi'} \lceil \bar{T}_{\pi'} / \lambda \rceil, \tag{7.2.14}$$

that is valid for $\operatorname{conv} \mathcal{R}$ cf. Günlük and Pochet [54, Theorem 2.1].

Remark 7.2.3. The important thing to note at this point is that the maximum right-hand sides $\bar{T}_\pi$ and $\bar{T}_{\pi'}$ for the two partitions may very well be attained for different scenarios. Hence the mixed partition inequality (7.2.14) may combine demand information from distinct scenarios and for this reason this class of inequalities may have greater significance when solving the stochastic program than what was experienced by Günlük [53] for the deterministic problem. To test this conjecture we performed a series of preliminary computational experiments. We used branch-and-cut to solve the problem AT13t (see Section 7.4) with 10 scenarios. Ten independent runs were performed with and without the mixed partition inequalities. These test runs revealed a significant reduction in the CPU time (approximately 59%) as well as a significant reduction in the number of nodes in the branching tree (approximately 57%) when the mixed partition inequalities were employed.

7.2.2 Facet-Defining Inequalities

As previously mentioned, sufficient conditions for partition inequalities and mixed-integer rounding inequalities to define facets of $\operatorname{conv}\mathcal{R}^s$ (or $\operatorname{conv}\mathcal{R}^s_{x,y}$) for some $s \in \{1,\dots,S\}$ have been given by various authors. One should note, though, that even if such a facet also defines a facet of $\bigcap_{s=1}^S \operatorname{conv}\mathcal{R}^s$ (or $\bigcap_{s=1}^S \operatorname{conv}\mathcal{R}^s_{x,y}$), it does not necessarily define a facet of $\operatorname{conv}\mathcal{R}$ (or $\operatorname{conv}\mathcal{R}_{x,y}$) since, in general, we have

$$\operatorname{conv}\mathcal{R} \subseteq \bigcap_{s=1}^{S} \operatorname{conv}\mathcal{R}^s$$

and consequently

$$\operatorname{conv}\mathcal{R}_{x,y} \subseteq \bigcap_{s=1}^{S} \operatorname{conv}\mathcal{R}^s_{x,y}$$

and these inclusions may be strict.

In this section we provide sufficient conditions for (7.2.8) and (7.2.9), respectively, to define facets of $\operatorname{conv}\mathcal{R}_{x,y}$. First, we note the following result.

Proposition 7.2.3. *$\operatorname{conv}\mathcal{R}_{x,y}$ and $\operatorname{conv}\mathcal{R}^s_{x,y}$, $s = 1,\dots,S$, are full-dimensional polyhedrons.*

Proof. We only show that $\operatorname{conv}\mathcal{R}_{x,y}$ is full-dimensional since the proof is exactly similar for the remaining sets. First note that $\operatorname{conv}\mathcal{R}_{x,y}$ is non-empty. Let $(\bar{x},\bar{y}) \in \operatorname{conv}\mathcal{R}_{x,y}$. Next, add to $(\bar{x},\bar{y})$ each of the $2|E|$ unit vectors. The $2|E|+1$ points obtained this way all belong to $\operatorname{conv}\mathcal{R}_{x,y}$ and they are affinely independent. $\square$

Recall that given a partition π of the node set, we let s^* denote a scenario for which the right-hand side of the partition inequality (7.2.6) is maximized, i.e.

$$s^* \in \arg\max_{1\le s\le S}\{\bar{T}^s_\pi\},$$

and we let $\bar{T}_\pi = \bar{T}^{s^*}_\pi$. We can now prove the following.

Proposition 7.2.4. *Consider a partition $\pi = \{V_1, \dots, V_l\}$ of the node set V. If the partition inequality (7.2.8) defines a facet of* $\operatorname{conv} \mathcal{R}^{s^*}_{x,y}$ *then it also defines a facet of* $\operatorname{conv} \mathcal{R}_{x,y}$.

Proof. Consider the two induced faces, $F = \{(x,y) \in \mathcal{R}_{x,y} \mid x(E_\pi) + \lambda y(E_\pi) = \bar{T}_\pi\}$ and $F^{s^*} = \{(x,y) \in \mathcal{R}^{s^*}_{x,y} \mid x(E_\pi) + \lambda y(E_\pi) = \bar{T}_\pi\}$. If F^{s^*} is a facet of $\operatorname{conv} \mathcal{R}^{s^*}_{x,y}$, we can find $2|E|$ affinely independent points $(x^1, y^1), \dots, (x^{2|E|}, y^{2|E|}) \in F^{s^*}$ cf. Proposition 7.2.3. Now consider the points given by

$$(\hat{x}^i_e, \hat{y}^i_e) = \begin{cases} (x^i_e, y^i_e) & \text{if } e \in E_\pi; \\ (x^i_e + M, y^i_e) & \text{otherwise,} \end{cases} \qquad i = 1, \dots, 2|E|,$$

where $M > 0$ is some large number. By the definition of $\bar{T}_\pi$ and s^* we see that for any scenario all of the solutions $(\hat{x}^i, \hat{y}^i)$, $i = 1, \dots, 2|E|$, allow a feasible routing of all demand across the cut E_π as well as all internal demand in each node set $V_1, \dots, V_l$. Therefore we must have $(\hat{x}^1, \hat{y}^1), \dots, (\hat{x}^{2|E|}, \hat{y}^{2|E|}) \in F$. Furthermore, these points are obtained by adding the same vector to each of the points $(x^1, y^1), \dots, (x^{2|E|}, y^{2|E|})$ and hence they are affinely independent. □

In the exact same way we may prove the following result.

Proposition 7.2.5. *Consider a partition $\pi = \{V_1, \dots, V_l\}$ of the node set V. If the mixed-integer rounding inequality (7.2.9) defines a facet of* $\operatorname{conv} \mathcal{R}^{s^*}_{x,y}$ *then it also defines a facet of* $\operatorname{conv} \mathcal{R}_{x,y}$.

Propositions 7.2.4 and 7.2.5 are useful to us, since they allow us to use conditions derived for the deterministic capacitated network design problem to identify facet-defining inequalities for the stochastic program.

7.3 Solution Procedure

Problem (7.2.2) is a large-scale mixed-integer programming problem and may be solved as such by standard software packages. However, as always when working with stochastic programming problems one should exploit the special structure of the problem and hence we will use the formulation (7.2.3)-(7.2.4). In this section we present a modified version of the L-shaped algorithm for stochastic linear programming problems, combining ordinary Benders decomposition with a branch-and-cut scheme. The L-shaped algorithm was discussed in some detail in Section 2.3.1. The seminal idea is to project the feasible region of the problem onto the space of discrete first-stage variables and hence the approach is closely related to that followed by Bienstock et al. [16] and Mirchandani [93]. The projection is built in a master problem by imposing different kinds of cuts. In addition to the well-known optimality cuts and feasibility cuts that are generated through the solution of subproblems, the procedure uses heuristically generated facet-defining inequalities as cutting planes in the master problem. As mentioned above, this approach is combined with a branch-and-cut scheme to solve the stochastic integer program.

Since $\operatorname{conv} \mathcal{R}_{x,y}$ is a convex polyhedron, the condition $(x, y) \in \operatorname{conv} \mathcal{R}_{x,y}$ may be replaced by a finite number of *feasibility cuts* corresponding to the facets of $\operatorname{conv} \mathcal{R}_{x,y}$. In general, however, we cannot identify all of these cuts due to the integer requirements on the first-stage variables. Still, we have shown that the integral metric inequalities (7.2.5), and in particular the partition inequalities (7.2.8), provide necessary conditions for feasibility of the second-stage problems and hence we will use these inequalities as feasibility cuts. The mixed-integer rounding inequalities (7.2.9) and the mixed partition inequalities (7.2.14) are not strictly necessary for second-stage feasibility, but we will use them as a sort of feasibility cuts since they do define facets of $\operatorname{conv} \mathcal{R}_{x,y}$ under certain conditions as previously pointed out.

Moreover, just as in the ordinary L-shaped algorithm, the convex and piecewise linear second-stage value functions (7.2.4) may be represented by a number of linear models, referred to as *optimality cuts*. To be specific, by linear programming duality we have for each scenario $s \in \{1, \ldots, S\}$ that

$$Q^s(x, y) = \max_{l \in \{1,\ldots,L^s\}} \left\{ \sum_{i \in V} \sum_{k \in K} D^s_{ik} \rho^l_{ik} - \sum_{\{i,j\} \in E} (\sigma^l_{ij} + \tau^l_{ij})(C_{ij} + x_{ij} + \lambda y_{ij}) \right\},$$

where $(\rho^l, \sigma^l, \tau^l)$, $l = 1, \ldots, L^s$ are the dual extreme points of the s'th second-stage problem. Hence for $s \in \{1, \ldots, S\}$ we may replace the second-stage value function Q^s by a single variable θ^s and the constraints

$$\theta^s \geq \sum_{i \in V} \sum_{k \in K} D^s_{ik} \rho^l_{ik} - \sum_{\{i,j\} \in E} (\sigma^l_{ij} + \tau^l_{ij})(C_{ij} + x_{ij} + \lambda y_{ij}) \qquad l = 1, \ldots, L^s.$$

Remark 7.3.1. Note that we have chosen to employ a multicut approach, imposing cuts on the individual second-stage value functions Q^s, $s = 1, \ldots, S$, rather than on the expected recourse function $\mathcal{Q} = \sum_{s=1}^{S} p^s Q^s$. This allows us to pass more detailed information to the master problem in each iteration cf. the discussion in Section 2.3.2.

The algorithm progresses by sequentially solving a master problem and adding feasibility cuts or optimality cuts that are violated at the current solution. Violated optimality cuts as well as violated metric inequalities are identified by solving the second-stage problems, thereby generating the needed dual extreme points and dual extreme rays. The separation problem for the integral metric inequalities, the partition inequalities, the mixed-integer rounding inequalities, and the mixed partition inequalities, on the other hand, is in general $\mathcal{NP}$-hard and we may have to resort to heuristics to identify violated cuts of these types — an issue to which we will return. By appropriately defining matrices $D = (D_1, D_2, d)$ and $E = (E_1, E_2, E_3, e)$ representing the feasibility cuts and optimality cuts that have been included, we may state the master problem as

$$\begin{aligned}
\min\ & ax + by + \sum_{s=1}^{S} p^s \theta^s \\
\text{s.t. } & D_1 x + D_2 y \geq d, \\
& E_1 x + E_2 y + E_3 \theta \geq e, \\
& x, y \in \mathbb{R}^{|E|}_+,\ \theta^1, \ldots, \theta^S \in \mathbb{R}.
\end{aligned} \tag{7.3.1}$$

Algorithm 7.1

Step 1 (*Initialization*) Set $\bar{z} = \infty$, choose an initial set of constraints represented by D and E, and let the list of open problems $\mathcal{L}$ consist of problem (7.3.1).

Step 2 (*Termination/Node selection*) If $\mathcal{L} = \emptyset$, stop; the solution that yielded the upper bound $\bar{z}$ is optimal. Otherwise, select and remove a problem P from $\mathcal{L}$.

Step 3 (*Solve master problem*) Solve the current master problem P and let $(x^P, y^P, \theta^{P,1}, \dots, \theta^{P,S})$ be an optimal solution vector with objective value z^P. If $z^P \geq \bar{z}$, return to Step 2. Otherwise,

(i) if (x^P, y^P) contains a fractional element, decide whether to proceed by cutting (go to Step 4) or branching (go to Step 5);

(ii) if (x^P, y^P) is integral, solve the second-stage problem (7.2.4) for all scenarios, let $\bar{z} = \min\{\bar{z},\ ax^P + by^P + \sum_{s=1}^S p^s Q^s(x^P, y^P)\}$, and remove from $\mathcal{L}$ all problems P' with $z^{P'} \geq \bar{z}$. If $\theta^{P,s} = Q^s(x^P, y^P)$ for $s = 1, \dots, S$, return to Step 2. Otherwise, go to Step 4.

Step 4 (*Cut generation*) Identify a number of cuts that are violated at the current solution $(x^P, y^P, \theta^{P,1}, \dots, \theta^{P,S})$, and augment D and E by appending the new rows to the appropriate matrix. Go to Step 3.

Step 5 (*Branching*) Select an edge $\{i, j\} \in E$ such that x_{ij}^P or y_{ij}^P is fractional, and construct two new problems P' and P'', obtained from P by adding respectively either the constraints $x_{ij} \leq \lfloor x_{ij}^P \rfloor$ and $x_{ij} \geq \lceil x_{ij}^P \rceil$, or the constraints $y_{ij} \leq \lfloor y_{ij}^P \rfloor$ and $y_{ij} \geq \lceil y_{ij}^P \rceil$. Let $z^{P'} = z^{P''} = z^P$ and add the two problems to $\mathcal{L}$. Go to Step 2.

Remark 7.3.2. Since only a finite number of different cuts can possibly be generated, it is straightforward to prove finite convergence of Algorithm 7.1 to an optimal solution of problem (7.2.2), if only branching occurs whenever no violated cuts can be identified.

Remark 7.3.3. Note that the general outline of Algorithm 7.1, and in particular the formulation of Step 3 (i), allows for several alternative practical implementations. As described in the following section, we chose to implement the algorithm so as to work simultaneously toward obtaining integral solutions of the master problem and building proper representations of the second-stage value functions by means of optimality cuts, i.e we chose to branch and cut simultaneously. Alternatively, one could choose to give higher priority to either of these objectives, i.e. to branch only when no more violated optimality cuts can be generated (cut-first-branch-second) or to generate optimality cuts only when the solution of the master problem is integral (branch-first-cut-second). We did not investigate the practicability of any of these alternatives, but we note that Albareda-Sambola, van der Vlerk, and Ferńandez [2], compared different versions of a similar algorithm for a class of stochastic generalized assignment problems, and concluded that a branch-and-cut scheme, such as that outlined in the following section, performed superior to either a branch-first-cut-second or a cut-first-branch-second scheme as discussed above. We also note, however, that the internet protocol network design problem, discussed in Chapter 10, is in fact solved by an algorithm that is similar in many ways to Algorithm 7.1, but for which the branch-first-cut-second scheme proved superior.

7.4 Computational Experiments

We implemented Algorithm 7.1 in C++ using procedures from the callable library of CPLEX 6.6 to solve the master- and subproblems. In this section we give a few implementational details before presenting results of our computational experiments.

7.4.1 Implementational Details

The branch-and-cut segment of the algorithm was implemented in compliance with the guidelines provided by Günlük [53], to which we refer for a detailed description of this part of the algorithm. At initialization we generate a large number of valid inequalities (partition inequalities, mixed-integer rounding inequalities, and mixed partition inequalities), and store these in a cutpool. At each node of the branching tree the formulation of the current master problem is strengthened using valid inequalities from the cutpool that are violated by the current solution. At most ten valid inequalities are added at a time whereupon the master problem is solved. This process is repeated until no more violated inequalities can be identified. Next, we try to generate new valid inequalities that are violated by the current solution, and in particular all second-stage problems are solved to potentially generate (integral) metric inequalities or optimality cuts. Finally, inequalities with large slacks (e.g. more than 10%) are removed from the master problem and stored in the cutpool to control the size of the current problem. The generation of new optimality cuts is not allowed to affect the decision whether to keep cutting or proceed by branching. Hence in Step 3 (i) we choose to proceed by branching whenever the current solution contains a fractional element, all second-stage problems are feasible, and no more valid inequalities that are violated by the current solution can be identified.

As previously discussed, feasibility cuts (metric inequalities) and optimality cuts are generated through the solution of second-stage problems. Integral metric inequalities, partition inequalities, mixed-integer rounding inequalities, and mixed partition inequalities, on the other hand, are generated heuristically using procedures described by Günlük [53] and Bienstock et al. [16]. The heuristic used to generate integral metric inequalities is executed every time a metric inequality is generated. This heuristic simply divides all coefficients of the metric inequality by the smallest positive coefficient. If the resulting coefficients are integral an integral metric inequality is obtained by rounding up the right-hand side. To generate partition inequalities we use two alternative heuristics, building partitions of the node set in different ways. One is used to generate cutset inequalities only, and works simply by randomly selecting a given number of nodes to form the set V_1. The other heuristic is used for general partition inequalities ($l \geq 2$). This heuristic first randomly selects one node for each node set $V_1, \ldots, V_l$ in the partition. The remaining nodes are assigned one at a time to the node sets so as to minimize the difference between the left- and right-hand side of the related partition inequality. Once a partition is built in this way, the left- and right-hand side of the corresponding inequality is calculated to check if the cut is violated. Whenever violated partition inequalities are generated it is very easy to generate the corresponding mixed-integer rounding inequalities and mixed partition inequalities. (Mixed partition inequalities were only generated at initialization by mixing all partition inequalities that were tight at the root node.)

7.4.2 Problem Instances

The computational experiments were performed on two real-life instances previously studied in Bienstock and Günlük [17] and Günlük [53]. The first instance is a network representing the Atlanta area, containing 15 nodes and 22 edges. Since we are primarily interested in long-term planning where uncertainty is more significant, we chose as starting point the instance exhibiting the largest increase in demand, referred to as AT13t in the previous studies. The second instance is a denser network with 16 nodes and 49 edges, representing the New York area. Again we chose the instance with largest demand increase as starting point (NY17t). In the second instance there is no cost of flow and no existing capacity in the network. Both instances have fully dense traffic matrices.

For each network we performed a series of experiments with varying number of scenarios. Scenarios were generated randomly assuming some uncertainty in the overall demand level captured in a parameter μ as well as some regional (node dependent) fluctuations captured in parameters ρ_i, $i \in V$. Hence for $i, k \in V$ and $s \in \{1, \dots, S\}$ the demand between nodes i and k under scenario s was calculated as

$$D_{ik}^s = \mu^s \rho_i^s \rho_k^s D_{ik},$$

where D_{ik} is demand between nodes i and k in the deterministic problem and the random parameters are sampled from uniform distributions,

$$\begin{aligned} \mu^s &\sim U(0.8, 1.2), \quad s = 1, \dots, S, \\ \rho_i^s &\sim U(0.9, 1.1), \quad s = 1, \dots, S,\ i \in V. \end{aligned}$$

7.4.3 Computational Results

For the Atlanta problem we first generated instances with 1, 5, 10, 50, 100, and 500 scenarios. For each number of scenarios we randomly generated ten independent instances and ran the algorithm. At termination we recorded the number of generated cuts, the number of cuts remaining in the master problem, the number of nodes in the branching tree, the lower and upper bound, and the CPU time spent by the procedure. CPU times are reported as minutes:seconds. The numbers reported in Table 7.1 are all averages over the ten independent runs.

Table 7.1: Atlanta Problems

S	Opt. cuts	Feas. cuts	Heur. cuts	Total cuts	Cuts in master	Nodes	Lower bound	Upper bound	Gap	CPU time
1	76	19	1895	1990	80	189	509068.0	509068.0	0%	0:10
5	111	21	1889	2021	94	302	547736.9	547736.9	0%	0:41
10	217	26	1870	2113	109	439	589801.1	589801.1	0%	1:39
50	321	45	1873	2239	170	372	625171.1	625171.1	0%	5:58
100	430	47	1907	2384	266	259	633071.8	633071.8	0%	8:03
500	1430	108	1926	3464	863	379	642090.4	642090.4	0%	60:22

First of all we note that the algorithm terminated with an optimal solution in every run performed on the Atlanta problems in this series of experiments. The large increase in the number of optimality cuts and the size of the master problem is only natural, since we chose to place disaggregate optimality cuts on the S second-stage value functions separately, and hence at least one active cut exists for each scenario. We also note that the number of cuts generated by the heuristics is fairly constant. This is due to the fact that the partition inequalities generated at initialization were identical (with regard to the left-hand side) irrespective of the number of scenarios. Hence the number of cuts generated at initialization is almost identical for all runs and these cuts constitute the major part of the heuristically generated cuts. (An average of about 1860 cuts were generated at initialization for these instances.) Finally, we observed occasional "extreme" runs requiring a very large number of nodes, whereas the major part of the runs terminated after a few hundred nodes or fewer had been investigated. This tendency became even clearer when we ran the algorithm with 1000 scenarios in which case one run did not terminate after more than ten hours of CPU time. Even in this situation, however, the algorithm is able to produce very good lower and upper bounds in a relatively short amount of time. Figure 7.1 shows the development of the lower and upper bound for this extreme run with 1000 scenarios.

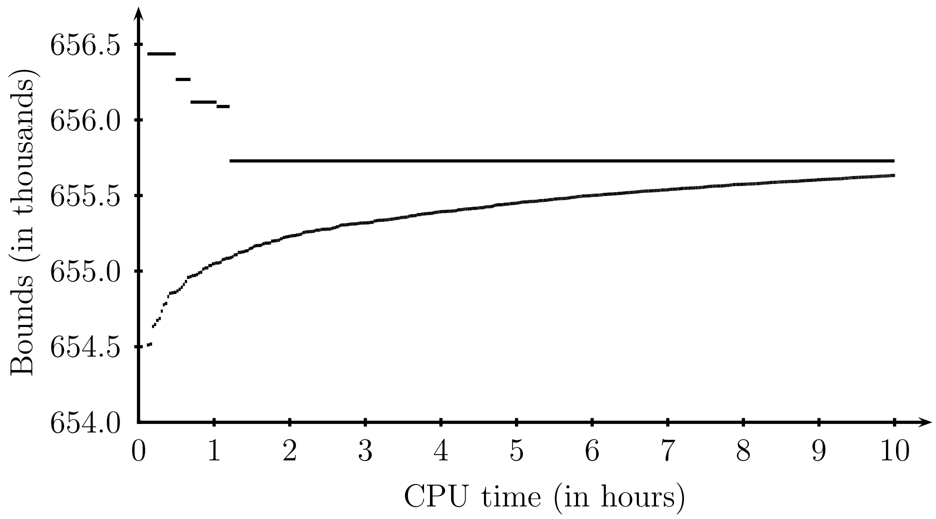

Figure 7.1: Atlanta Problem, 1000 Scenarios

We note that an initial lower bound of 650,529.7 and an initial upper bound of 658,651.6, each maintained during the first seven minutes of computation, are not shown in Figure 7.1 to facilitate proper scaling. From the figure we see that the gap between the lower and upper bound is quickly narrowed. Hence, the initial gap of 1.23% is narrowed to 0.29% after seven minutes of CPU time. After approximately one hour of CPU time the best (optimal?) solution is found and the gap is narrowed to 0.10%. Finally, the remaining gap after ten hours of CPU time is as small as 0.02%. We conclude that even with 1000 scenarios the algorithm usually terminated within a few hours of CPU time and when this was not the case, very good upper and lower bounds were provided in a reasonable amount of time.

Next, we turn to the New York problems. These problems are much harder than the Atlanta problems and the algorithm required almost two hours of CPU time to solve the deterministic problem NY17t. Hence, our main interest lay in the quality of the bounds provided in a reasonable amount of time, and we chose to stop all runs after three hours of CPU time. Since there is no cost of flow in these problems, we did not place any optimality cuts. Because all runs, performed on instances including multiple scenarios, terminated with a non-optimal solution, we report not only the average gap between the lower and upper bound but also the maximum gap from the ten independent runs. Apart from this the statistics appearing in Table 7.2 are the same as those recorded for the Atlanta problems.

Table 7.2: New York Problems

S	Feas. cuts	Heur. cuts	Total cuts	Cuts in master	Nodes	Lower bound	Upper bound	Gap	Max gap	CPU time
1	11118	18341	29459	135	6931	3780.8	3780.8	0.00%	0.00%	108:09
5	5200	17837	23037	189	3156	4214.2	4268.8	1.27%	2.14%	180:00
10	3250	17664	20914	197	1545	4295.4	4366.6	1.63%	2.14%	180:00
50	874	17585	18459	250	399	4482.2	4584.3	2.23%	3.54%	180:00
100	410	17597	18007	246	254	4514.9	4626.6	2.41%	3.76%	180:00
500	207	17688	17895	216	47	4566.9	4752.0	3.89%	4.88%	180:00

First of all we note the quality of the bounds provided by the algorithm. Thus, we see that the average gap as well as the maximum gap is modest for all six series of experiments, even though a significant increase in the gap is observed when the number of scenarios increase. Naturally, the increased gap is caused by the drastic decrease in the number of feasibility cuts and the number of nodes investigated which result from the increased computation time required per iteration when the number of scenarios increase. On the other hand we see that the number of heuristically generated cuts is once again fairly constant due to the large number of cuts generated at initialization. (An average of about 17,500 cuts were generated at initialization for these instances.) The number of cuts remaining in the master problem at termination does not exhibit the same sensitivity with respect to the number of scenarios as for the Atlanta problems, since no optimality cuts were placed.

Acknowledgment

The author would like to thank Oktay Günlük for providing the problem instances for computational testing.

Chapter 8

A Bicriterion Model for Capacity Expansion

In this chapter we consider once again the capacity expansion of a telecommunications network in the face of uncertainty. Here, the uncertainties facing the network operator are assumed to be twofold. Some arises due to the inherent uncertainty involved in the assessment of future demand, and some is due to the potential future failure of nodes or edges in the network. Previous studies have mainly dealt with these issues separately. Apart from the work presented in this thesis, capacity expansion of telecommunication networks with uncertain future demand has been considered by authors such as Dempster, Medova, and Thompson [40], Medova [90], and Sen, Doverspike, and Cosares [145]. Capacity expansion problems with potential future failure of network components, on the other hand, have been considered in the framework of survivable network design by e.g. Dahl and Stoer [36, 149], Grötschel, Monma, and Stoer [52], Minoux [92], and Rios, Marianov, and Gutierrez [120]. All of these capacity expansion models actually fit into the general framework of two-stage stochastic programs with recourse, the first stage usually corresponding to the planning of capacity installments, and the second stage usually corresponding to the routing of traffic in the network once actual demand has been observed and a failure has possibly occurred. We note, however, that this terminology is usually not used in connection with survivable network design. In this chapter, the capacity expansion problem is formulated as a bicriterion stochastic program with recourse in which the probability of future capacity requirements to be violated in case of network failures and the total cost of the capacity expansion are simultaneously minimized.

8.1 Problem Formulation

The network is modeled as a connected undirected graph $G = (V, E)$, where V denotes the set of nodes (switches) and E denotes the set of edges (circuit groups). As in Chapter 7, demand is described by a set K of commodities. In the present context, however, we will find a disaggregated formulation of the commodities most convenient, and hence we choose to let each commodity $k \in K$ correspond to a point-to-point pair of nodes between which demand is to be routed. Furthermore, as in the preceding chapters, uncertainty is incorporated in the formulation by introducing a finite number of scenarios representing the possible future states of the world. Associated with a scenario $s \in \{1, \dots, S\}$ is a specific failure state (possibly no failure) reducing the set of functional nodes and edges to V^s and E^s, respectively, and a set of point-to-point demands D_k^s to be routed between functional node-pairs $k \in K^s$. We will assume that the routing of traffic is restricted to a set of prespecified routes $\mathcal{P}$. Given a scenario $s \in \{1, \dots, S\}$, the set of functional routes between a node-pair $k \in K^s$ is denoted by $\mathcal{P}_k^s$ and the set of functional routes which use the edge $\{i, j\} \in E^s$ is denoted by $\mathcal{Q}_{ij}^s$. Also, for $s \in \{1, \dots, S\}$ we let $\mathcal{P}^s = \bigcup_{k \in K^s} \mathcal{P}_k^s$ denote the set of all functional routes under scenario s. For $s \in \{1, \dots, S\}$ we assume that the probability of scenario s to actually occur is known and we denote it by π^s. The existing capacity on an edge $\{i, j\} \in E$ is denoted by C_{ij}. Additional capacity may be installed in multiples of a fixed batch size. In particular, we assume that a single facility, providing a capacity of λ, is available for installation on all edges $\{i, j\} \in E$ at a unit cost of c_{ij}. By rescaling demand and existing capacity, we may assume that $\lambda = 1$.

Along the lines of Remark 7.2.1 on page 96, we note that the demand input for the long-term network planning model under consideration here is a set of capacity requirements between node-pairs, required to maintain some prescribed grade-of-service (GoS). (See also Dempster, Medova, and Thompson [40] and Medova [90] for related discussions.) In the case of node or edge failures it is required that a certain fraction of these capacity requirements is available to uphold the GoS for all node-pairs. Since the prescribed GoS, as well as the fraction of capacity requirements to be available in case of failures, are selected somewhat arbitrarily, however, refusing to waive these requirements under any circumstances may not make sense in a cost minimization framework. In other words, we may obtain a considerable decrease in the optimal cost by relaxing the requirements for a few critical failure states. Moreover, such a cost reduction would be of major interest if the probability of the critical failures to actually supervene is very small. To illuminate the trade-off between total cost of the capacity expansion and the probability of GoS requirements to be violated, we formulate a bicriterion model for capacity expansion in which these two objectives are simultaneously minimized. To this end, we denote by x_{ij} the number of facilities to be installed on edge $\{i, j\} \in E$, and introduce for $s \in \{1, \dots, S\}$ an indicator function, $\psi^s : \mathbb{R}_+^{|E|} \mapsto \{0, 1\}$, such that $\psi^s(x) = 1$ if and only if the capacity installment $x \in \mathbb{R}_+^{|E|}$ is not sufficient to meet the capacity requirements under scenario s. The bicriterion stochastic programming model may now be formulated as

$$\begin{aligned} \min\ z_1 &= \sum_{\{i,j\} \in E} c_{ij} x_{ij} \\ \min\ z_2 &= \sum_{s=1}^{S} \pi^s \psi^s(x) \\ \text{s.t. } & x \in \mathbb{Z}_+^{|E|}, \end{aligned} \tag{8.1.1}$$

Here the first objective z_1 is the total cost of the capacity expansion whereas the second objective z_2 is the probability of capacity requirements to be violated. For practical purposes we find it convenient for $s \in \{1, \dots, S\}$ to define the indicator function ψ^s by

$$\psi^s(x) = \begin{cases} 1 & \text{if } \phi^s(x) > 0; \\ 0 & \text{otherwise,} \end{cases} \tag{8.1.2}$$

where the function $\phi^s : \mathbb{R}_+^{|E|} \mapsto \mathbb{R}$ is given by

$$\begin{aligned} \phi^s(x) = \min & \sum_{k \in K^s} t_k^s \\ \text{s.t.} & \sum_{p \in \mathcal{P}_k^s} f_p^s + t_k^s \geq \rho_k^s D_k^s, \quad && k \in K^s, \\ & \sum_{p \in \mathcal{Q}_{ij}^s} f_p^s \leq C_{ij} + x_{ij}, \quad && \{i,j\} \in E^s, \\ & f_p^s,\, t_k^s \geq 0, && p \in \mathcal{P}^s,\ k \in K^s. \end{aligned} \tag{8.1.3}$$

Here, for $s \in \{1, \ldots, S\}$ and $p \in \mathcal{P}^s$ the variable f_p^s denotes the amount of capacity to be allocated to route p under scenario s. Moreover, for $s \in \{1, \ldots, S\}$ and $k \in K^s$ the parameter ρ_k^s denotes the minimum fraction of the capacity requirement between node-pair k that should be available under scenario s, whereas the variable t_k^s denotes the corresponding shortage of capacity between node-pair k under scenario s, which is to be minimized. Hence we see that (8.1.2)-(8.1.3) provides an appropriate definition of the indicator function.

Remark 8.1.1. Modeling the actual process of real-time call-by-call routing within a long-term planning model as the one considered here, is obviously not viable. As briefly pointed out above, the demand input for this problem is a set of capacity requirements between point-to-point pairs of nodes, needed to maintain a prescribed grade-of-service. For example, as discussed in Remark 7.2.1 on page 96, Dempster, Medova, and Thompson [40] and Medova [90] determine these capacity requirements as the effective bandwidth requirements needed to ensure that a set of blocking probabilities are not exceeded. In Sen, Doverspike, and Cosares [145] the approximation of real-time routing by a static model similar to problem (8.1.3) was validated using simulation with encouraging results.

Remark 8.1.2. The assumption that routing of traffic is restricted to a set of prespecified routes is a common one, employed also by e.g. Dempster, Medova, and Thompson [40], Medova [90], and Sen, Doverspike, and Cosares [145]. The assumption may be justified by the fact that most static real-time routing algorithms implemented in switch software choose routes from a limited set, allowing us to simply enumerate the routes of interest.

Remark 8.1.3. As pointed out above, the capacity expansion problem (8.1.1) fits in the general framework of two-stage stochastic recourse programs. The first stage includes the decisions on capacity expansion which must be made before the future state of the world is known. Once uncertainty is revealed, the second-stage decision, consisting of allocation of capacity to routes, is settled. One might argue, though, that a three-stage formulation of the problem would more accurately capture the actual alternating process of decisions and observations of random outcomes. Here, as before, the decision on capacity expansion is made in the first stage. In the second stage, an actual outcome of random demand is observed and the capacity is allocated accordingly to routes. Finally, in the third stage, a failure possibly occurs and capacity may be reallocated among routes. Assuming that the second-stage allocation of capacity to routes is of no importance for the reallocation of capacity in the third stage, however, the second and third stages may in fact be joined to obtain the two-stage formulation (8.1.1).

Remark 8.1.4. We note the close resemblance between problem (8.1.1) and the mean-risk models that were discussed in Section 4.1.1 in connection with the minimum risk problem. In particular, we recall that the probability-based objective function was studied in detail throughout Chapter 4, where structural properties were investigated and, in particular, we established certain stability results that justify the approach of representing uncertainty by means of a finite number of scenarios. In Chapter 4 we also elaborated an algorithm for the minimum risk problem, and the seminal idea of this approach is a corner stone in the solution procedure for problem (8.1.1) presented in the following section.

8.2 Solution Procedure

Before we proceed let us consider for a moment a general bicriterion problem of the form

$$\begin{aligned} &\min\ z_1 = f_1(x) \\ &\min\ z_2 = f_2(x) \\ &\ \text{s.t. } x \in X. \end{aligned} \tag{8.2.1}$$

Since, in general, we cannot expect to obtain a solution $\bar{x} \in X$ that minimizes both objectives over X, it is not immediately clear what an "optimal" solution of problem (8.2.1) should be. The relevant concept in this respect is that of efficient solutions, defined next. Let the feasible region in criterion space be

$$\mathcal{Z} = \big\{(z_1, z_2) \in \mathbb{R}^2 \,\big|\, \exists\, x \in X : z_1 = f_1(x),\ z_2 = f_2(x)\big\}.$$

Definition 8.2.1. A criterion vector $(z_1, z_2) \in \mathcal{Z}$ is dominated if there exists $x \in X$ such that $f_1(x) \le z_1$ and $f_2(x) \le z_2$ with at least one inequality being strict. Otherwise (z_1, z_2) is a non-dominated criterion vector.

Definition 8.2.2. A solution vector $x \in X$ is efficient if the corresponding criterion vector $\big(f_1(x), f_2(x)\big)$ is non-dominated. Otherwise x is inefficient.

For a basic introduction to multiple criteria optimization we refer to Steuer [148].

8.2.1 Finding all Non-Dominated Solutions

To determine all non-dominated solutions of problem (8.1.1), we observe that the second objective z_2 can only take on a finite number of values, say $p_1, \dots, p_n$, in any solution to the problem. Hence to obtain all non-dominated solutions to the bicriterion problem, we may simply solve the following problem for all such possible values p,

$$\begin{aligned} \min\ z_1 &= \sum_{\{i,j\} \in E} c_{ij} x_{ij} \\ \text{s.t. } \ z_2 &= \sum_{s=1}^{S} \pi^s \psi^s(x) \le p, \\ x &\in \mathbb{Z}_+^{|E|}. \end{aligned} \tag{8.2.2}$$

We will refer to problem (8.2.2) as the p-restricted problem.

Algorithm 8.1

Step 1 (*Initialization*) Let $0 = p_1 < \cdots < p_n$ be the possible values of $\sum_{s\in\mathcal{S}} \pi^s$, $\mathcal{S} \subseteq \{1, \ldots, S\}$, to be considered. Set $z_1^0 = \infty$, $\mathcal{L} = \emptyset$, and $i = 1$.

Step 2 (*Solve problem*) Solve the p-restricted problem (8.2.2) with $p = p_i$ and let (x^i, z_1^i, z_2^i) be an optimal solution vector.

Step 3 (*Update list*) If $z_1^i < z_1^{i-1}$ then set $\mathcal{L} = \mathcal{L} \cup \{(x^i, z_1^i, z_2^i)\}$.

Step 4 (*Termination*) If $i = n$ then stop. Otherwise set $i = i + 1$ and go to Step 2.

Remark 8.2.1. Note that problem (8.2.2) is feasible for all possible values $p \in [0, 1]$, since z_2 can be made arbitrarily small (equal to zero) by installing sufficient capacity. Hence an optimal solution can always be found in Step 2 of Algorithm 8.1.

Remark 8.2.2. In general the number of p-restricted problems to be solved during the course of Algorithm 8.1 may be very large ($n \leq 2^S$). For all practical purposes, however, it will often be sufficient to consider only a modest number of possible values for p. This happens for two primary reasons. First of all the network operator is most likely to accept only very small values of p and hence all values p_i, $i \in \{1, \ldots, n\}$, that exceed some maximum acceptable level may be discarded beforehand. Secondly, the number of possible values is reduced if any of the scenario probabilities are equal. In particular, if we employ a uniform probability distribution, i.e. $\pi^1 = \cdots = \pi^S$, then we have $n \leq S + 1$. Uniform scenario probabilities are often used in practical studies, for example when scenarios are generated by sampling.

Proposition 8.2.1. *At termination of Algorithm 8.1, the set of non-dominated criterion vectors is* $\{(z_1, z_2) \in \mathbb{R}^2 \mid \exists\, x \in \mathbb{Z}_+^{|E|} : (x, z_1, z_2) \in \mathcal{L}\}$.

Proof. First note that $z_1^1 \geq \cdots \geq z_1^n$ since $p_1 < \cdots < p_n$. Now, assume that in some iteration i of the algorithm we have $z_1^i < z_1^{i-1}$ but (z_1^i, z_2^i) is a dominated criterion vector, i.e. there exists a solution (x, z_1, z_2) to problem (8.1.1) such that $z_1 \leq z_1^i$ and $z_2 \leq z_2^i$ with at least one inequality being strict. Now, we must have $z_2 < z_2^i$, since $z_1 < z_1^i$ contradicts optimality of (x^i, z_1^i, z_2^i) in problem (8.2.2) with $p = p_i$. Note that this may only happen for $i \geq 2$ because $p_1 = 0$, and hence it follows that $z_2 = p_j$ for some $j \in \{1, \ldots, i-1\}$. But this contradicts optimality of (x^j, z_1^j, z_2^j) in problem (8.2.2) with $p = p_j$ since we have $z_1^j \geq z_1^{i-1} > z_1^i = z_1$ by assumption. Hence we see that only solutions for which the criterion vector is non-dominated are put into the list $\mathcal{L}$.

To see that $\{(z_1, z_2) \in \mathbb{R}^2 \mid \exists\, x \in \mathbb{Z}_+^{|E|} : (x, z_1, z_2) \in \mathcal{L}\}$ contains all non-dominated criterion vectors, assume that in some iteration i of the algorithm, the solution (x^i, z_1^i, z_2^i) is not put into the list $\mathcal{L}$, i.e. we have $z_1^i = z_1^{i-1}$. (Note that this can only happen for $i \geq 2$). If $z_2^i > z_2^{i-1}$ the criterion vector (z_1^i, z_2^i) is dominated by (z_1^{i-1}, z_2^{i-1}). So assume on the contrary that $z_2^i \leq z_2^{i-1}$ so that $z_2^i = p_j$ for some $j \in \{1, \ldots, i-1\}$. Now, we obviously have that $z_2^j \leq z_2^i$, and since (x^i, z_1^i, z_2^i) is a feasible solution for problem (8.2.2) with $p = p_j$ we must also have $z_1^j \leq z_1^i$. Thus the criterion vector (z_1^i, z_2^i) is either equal to or dominated by (z_1^j, z_2^j) and hence the solution (x^i, z_1^i, z_2^i) can be excluded from the list $\mathcal{L}$ with no loss of non-dominated solutions. The result follows, since all possible values of z_2 are considered during the course of the algorithm. ∎

Remark 8.2.3. Clearly, it may happen that several efficient solutions correspond to the same non-dominated criterion vector. Hence if $\{x \in \mathbb{Z}_+^{|E|} \mid \exists\, z_1, z_2 \in \mathbb{R} : (x, z_1, z_2) \in \mathcal{L}\}$ is required to contain all efficient solution vectors, it is necessary in Step 2 of Algorithm 8.1 to determine *all* optimal solutions of problem (8.2.2) in each iteration i where $z_1^i < z_1^{i-1}$.

8.2.2 Solving the p-Restricted Problems

The question remains how to efficiently solve problem (8.2.2) in Step 2 of Algorithm 8.1. The approach we suggest here is based on the seminal idea presented in Chapter 4 in connection with the solution procedure for the minimum risk problem. We recall that the idea is for each scenario $s \in \{1, \ldots, S\}$ to replace the indicator function ψ^s by a binary variable and a number of cutting planes derived through linear programming duality. In particular, for any feasible solution $x \in \mathbb{Z}_+^{|E|}$ and for any scenario $s \in \{1, \ldots, S\}$, the binary variable θ^s representing $\psi^s(\cdot)$ at x should be equal to one if and only if $\phi^s(x) > 0$, where the function ϕ^s was defined by (8.1.3). Consider now the dual of problem (8.1.3) for some $s \in \{1, \ldots, S\}$. We let $M > 0$ be some upper bound on the optimal value of this problem and we denote by $\mathcal{D}^s$ the set of extreme points of the feasible region. From the discussion above, we immediately have the following results.

Lemma 8.2.1. *For all $x \in \mathbb{R}_+^{|E|}$ and for any scenario $s \in \{1, \ldots. S\}$, the indicator function $\psi^s(x)$ satisfies the following set of inequalities,*

$$M\psi^s(x) \geq \sum_{k \in K^s} \rho_k^s D_k^s u_k - \sum_{\{i,j\} \in E^s} (C_{ij} + x_{ij}) v_{ij}, \quad (u, v) \in \mathcal{D}^s.$$

Lemma 8.2.2. *Let $x \in \mathbb{R}_+^{|E|}$ be such that $\psi^s(x) = 1$ for some scenario $s \in \{1, \ldots. S\}$. Then there exists $(u, v) \in \mathcal{D}^s$ such that*

$$\sum_{k \in K^s} \rho_k^s D_k^s u_k - \sum_{\{i,j\} \in E^s} (C_{ij} + x_{ij}) v_{ij} > 0.$$

Lemmas 8.2.1 and 8.2.2 elucidate the structure of the cutting planes, and in particular we see that the p-restricted problem (8.2.2) is equivalent to the following problem,

$$\begin{aligned}
\min \quad & \sum_{\{i,j\} \in E} c_{ij} x_{ij} \\
\text{s.t.} \quad & \sum_{s=1}^{S} \pi^s \theta^s \leq p, \\
& \sum_{k \in K^s} \rho_k^s D_k^s u_k - \sum_{\{i,j\} \in E^s} (C_{ij} + x_{ij}) v_{ij} \leq M\theta^s, \quad (u, v) \in \mathcal{D}^s,\ s = 1, \ldots, S, \\
& x \in \mathbb{Z}_+^{|E|},\ \theta^1, \ldots, \theta^S \in \{0, 1\}.
\end{aligned}$$

Remark 8.2.4. The cutting planes described above may be considered as generalizations of the metric inequalities originally introduced by Iri [63] and Onaga and Kakusho [104]. These inequalities were discussed in detail in Section 7.1.1 in connection with the capacitated network design problem, where they were employed as valid inequalities in the cutting plane procedure elaborated for the problem.

Remark 8.2.5. We note that problem (8.1.3) is always feasible and bounded and hence the same things go for its dual. Thus, optimal solutions to the problems always exist, and their optimal values are equal. Moreover, an upper bound M on the optimal value of the problems, obtained by letting $f_p^s = 0$ for all $p \in \mathcal{P}^s$ and $s \in \{1, \dots, S\}$, is conveniently calculated as

$$M = \max_{s \in \{1,\dots,S\}} \left\{ \sum_{k \in K^s} \rho_k^s D_k^s \right\}.$$

The algorithm progresses by sequentially solving a master problem and adding violated cutting planes generated through the solution of subproblems (8.1.3). Hence for some subsets $\mathcal{E}^s \subseteq \mathcal{D}^s$, $s = 1, \dots, S$, of the dual extreme points, we define the master problem as the following relaxation in which only some of the cutting planes are included,

$$\begin{aligned}
\min\ & \sum_{\{i,j\} \in E} c_{ij} x_{ij} \\
\text{s.t. } & \sum_{s=1}^{S} \pi^s \theta^s \leq p, \\
& \sum_{k \in K^s} \rho_k^s D_k^s u_k - \sum_{\{i,j\} \in E^s} (C_{ij} + x_{ij}) v_{ij} \leq M \theta^s, \quad (u, v) \in \mathcal{E}^s,\ s = 1, \dots, S, \\
& x \in \mathbb{Z}_+^{|E|},\ \theta^1, \dots, \theta^S \in \{0, 1\}.
\end{aligned} \tag{8.2.3}$$

Algorithm 8.2

Step 1 (*Initialization*) Set $\nu = 0$, for $s = 1, \dots, S$ let $\mathcal{E}^s \subseteq \mathcal{D}^s$, and let the initial master problem be defined by (8.2.3).

Step 2 (*Solve master problem*) Set $\nu = \nu + 1$. Solve the current master problem and let $(x^\nu, \theta^{1,\nu}, \dots, \theta^{S,\nu})$ be an optimal solution vector.

Step 3 (*Solve subproblems*) Solve the second-stage problem (8.1.3) for all scenarios $s \in \{1, \dots, S\}$ such that $\theta^{s,\nu} = 0$. Consider the following situations,

(i) if $\phi^s(x^\nu) = 0$ for all of these scenarios, stop — the current solution x^ν is optimal for the p-restricted problem (8.2.2);

(ii) if $\phi^s(x^\nu) > 0$ for some of these scenarios, say for $s \in \mathcal{S} \subseteq \{1, \dots, S\}$, then a dual extreme point $(u^s, v^s) \in \mathcal{D}^s$ with positive objective value is identified for each $s \in \mathcal{S}$, and the corresponding cutting planes are added to the master problem. Set $\mathcal{E}^s = \mathcal{E}^s \cup \big\{(u^s, v^s)\big\}$ for each $s \in \mathcal{S}$ and return to Step 2.

Remark 8.2.6. Recall that Algorithm 8.1 involved the solution of problem (8.2.2) for a sequence of increasing values of p. The cutting planes generated while solving the first of these problems remain valid when p is changed. Hence in Step 1 of Algorithm 8.2, we may let the sets $\mathcal{E}^s$, $s = 1, \dots, S$, consist of the dual extreme points generated in previous runs of Algorithm 8.2 (or some subset thereof). This strategy of retaining cutting planes from previous runs resulted in remarkable time savings in the overall solution time for problem (8.1.1).

Proposition 8.2.2. *Algorithm 8.2 terminates with an optimal solution of the p-restricted problem (8.2.2) in a finite number of iterations.*

Proof. First of all note that an optimal solution of the p-restricted problem (8.2.2) always exists cf. Remark 8.2.1. Second, note that the optimal value of the master problem in any iteration is a lower bound on the optimal value of problem (8.2.2), since the master problem is a relaxation. Now, suppose that in some iteration ν and for some scenario $s \in \{1, \dots, S\}$ we have $\theta^{s,\nu} < \psi^s(x^\nu)$. In that case a violated cutting plane, cutting off the current solution $(x^\nu, \theta^{1,\nu}, \dots, \theta^{S,\nu})$, is identified in Step 3 cf. Lemma 8.2.2, and the algorithm proceeds. Since the number of dual extreme points is finite, this can only happen a finite number of times, and we will eventually have $\theta^{s,\nu} \geq \psi^s(x^\nu)$ for all $s \in \{1, \dots, S\}$, and hence

$$\sum_{s=1}^{S} \pi^s \psi^s(x^\nu) \leq \sum_{s=1}^{S} \pi^s \theta^{s,\nu} \leq p.$$

At this point the current solution x^ν is feasible in problem (8.2.2) and hence optimal. The corresponding criterion vector is $(z_1, z_2) = \left(\sum_{\{i,j\} \in E} c_{ij} x_{ij}^\nu \,, \sum_{s=1}^{S} \pi^s \theta^{s,\nu}\right)$. □

Remark 8.2.7. MirHassani et al. [94] considered a capacity expansion problem arising in supply chain network planning and solved the problem by a Benders decomposition approach, similar in many ways to Algorithm 8.2. The authors observed that solutions to the master problem in early iterations performed very poorly, since the master problem tends to minimize the amount of capacity installed, whereas "good" solutions in the second stage require substantial amounts of capacity to be installed. To circumvent this problem, several enhancements of the master problem were considered. One such enhancement was to include some scenario in the master problem, thus making it more representative of the second-stage subproblems. Using this expanded formulation MirHassani et al. observed a considerable improvement in the overall solution time. In our setting, on the other hand, such an expanded formulation performed very poorly. This is not too surprising, though, since the expanded master problem in this case is a capacitated network design problem with additional constraints, and several studies have shown that projecting out the flow variables is an efficient solution approach for such problems cf. Chapter 7 and references therein. Also, when the above-mentioned strategy of retaining cuts from previous runs is employed, the lack of consistency between the master problem and the second-stage subproblems is only significant in early iterations of the first run ($p = 0$) and hence does not outweigh the increased effort required to solve an expanded master problem.

8.2.3 Valid Inequalities for the p-Restricted Problems

The algorithm for the p-restricted problem (8.2.2), proposed in the previous section, solves a sequence of master problems which are all integer programming problems. In early iterations, though, one should not put too much effort into finding optimal integer solutions for these problems, since solutions are cut off anyway, as more cutting planes are added. Hence, rather than solving the integer master problem (8.2.3) to optimality in each iteration, we chose to work with a relaxation of the problem and strengthen the formulation using valid inequalities.

From now on we will refer by the linear relaxation of the master problem (8.2.3) to the corresponding problem in which integer requirements on the capacity variables have been relaxed. (Hence we speak of a linear relaxation, even though we still have binary variables $\theta^1, \dots, \theta^S$.) Starting from this relaxation, we add cutting planes defining the indicator functions as described in the previous section. These cutting planes, however, should not only be used to define the indicator functions, but also as valid inequalities for the convex hull of feasible integer solutions. In particular, since the feasible region of the dual of problem (8.1.3) is a rational polyhedron, we may assume that the extreme points $(u, v) \in \mathcal{D}^s$, $s = 1, \dots, S$, are integral — this can be achieved by scaling. Hence the cutting planes derived in the previous section may be strengthened by rounding. Applying this approach, we arrive at what we will refer to as the strengthened linear relaxation of the master problem,

$$
\begin{aligned}
\min \; & \sum_{\{i,j\} \in E} c_{ij} x_{ij} \\
\text{s.t.} \; & \sum_{s=1}^{S} \pi^s \theta^s \leq p, \\
& \sum_{\{i,j\} \in E^s} x_{ij} v_{ij} + M\theta^s \geq \Big\lceil \sum_{k \in K^s} \rho_k^s D_k^s u_k - \sum_{\{i,j\} \in E^s} C_{ij} v_{ij} \Big\rceil, \\
& \qquad\qquad\qquad\qquad\qquad\qquad (u, v) \in \mathcal{E}^s, \; s = 1, \dots, S, \\
& x \in \mathbb{R}_+^{|E|}, \; \theta^1, \dots, \theta^S \in \{0, 1\}.
\end{aligned}
\tag{8.2.4}
$$

Obviously, solving in each iteration problem (8.2.4) rather than problem (8.2.3) may not produce an integer solution at termination of Algorithm 8.2. Therefore, to obtain an optimal integer solution of the p-restricted problem (8.2.2), this approach should be combined with some appropriate branching scheme. To this end, clearly, a possible approach is simply to explicitly reintroduce the integer requirements on capacity variables in the master problem and proceed with Algorithm 8.2. Alternatively, the solution procedure could be incorporated in a more extensive branch-and-cut scheme similar to the one elaborated in Chapter 7 for the capacitated network design problem. As described in the following section, though, we only implemented the former alternative. Furthermore, as the computational experiments presented in Section 8.3 indicate, the integrality gap is very small when the strengthened linear relaxation of the master problem is put to use. Therefore we also consider it a viable approach to restrict attention to the strengthened linear relaxation of the master problem and subsequently establish near-optimal integer solutions by some suitable heuristic.

Remark 8.2.8. Along the lines of Remark 8.2.4, we note that the cutting planes used in the strengthened linear relaxation of the master problem (8.2.4) may be seen as a generalization of the class of integral metric inequalities which are valid inequalities for the capacitated network design problem, discussed in Chapter 7. In a similar manner one may derive generalizations of other classes of valid inequalities discussed in connection with the capacitated network design problem such as e.g. partition inequalities and mixed-integer rounding inequalities.

8.3 Computational Experiments

The solution procedure for the bicriterion capacity expansion problem (8.1.1), described in the previous section, was implemented in C++ using procedures from the callable library of CPLEX 6.6 to solve the linear subproblems (8.1.3) and the (mixed-) integer master problems (8.2.3) or (8.2.4). A series of computational experiments were carried out to test the quality of the approximations provided by the strengthened linear relaxation and an upper bounding heuristic, and to investigate the practicability of the algorithm.

8.3.1 Implementational Details

As briefly discussed in Section 8.2.3 we chose to relax the integer requirements on x in the master problem at initialization of each run. Starting from this relaxation we proceeded with Algorithm 8.2 until no more cuts could be identified. If the current solution at this point was not integral we explicitly reintroduced the integer requirements on x in the master problem and proceeded with Algorithm 8.2 until no more cuts could be identified. Using this approach it turned out that, in general, only very few additional iterations upon reintroduction of the integer requirements on x were necessary before an optimal integer solution of the current p-restricted problem was achieved. Hence we did not find it worthwhile to generate cuts during branching in a more extensive branch-and-cut scheme such as that elaborated for the capacitated network design problem in Chapter 7.

Cutting planes for the master problem were generated through the solution of subproblems (8.1.3). To obtain the generalized integral metric inequalities described in Section 8.2.3 we used a heuristic discussed also in Section 7.4.1. This heuristic simply divides all coefficients of the cut by the smallest positive coefficient. If the resulting coefficients are integral, a generalized integral metric inequality is obtained by rounding up the right-hand side.

As pointed out in Section 8.2.2, we achieved a considerable reduction in overall solution time by keeping cuts from previous runs in the master problem when the value of p was updated. To control the size of the master problem, however, it was necessary to temporarily remove "old" cuts. For $s \in \{1, \ldots, S\}$, a cut $ax + M\theta^s \geq b$ was considered to be inactive if the corresponding binary variable θ^s was equal to 0 in the current solution and the relative slack $(ax - b)/b$ was larger than 10%. A cut which had been inactive for more than 10 iterations was temporarily removed from the master problem and stored in a cutpool. The cutpool, on the other hand, was searched at regular intervals, and any violated cuts were returned to the master problem. The definition of inactive cuts and the number of iterations to keep an inactive cut in the master problem were chosen somewhat arbitrarily, so as to keep the size of the master problem manageable, while limiting the number of movements in and out of the cutpool.

Even though the additional number of iterations required upon reintroduction of the integer requirements on x in the master problem was small, the CPU time required for these additional iterations turned out to be substantial, at least for the larger instances. Hence for some problems it may not be practicable to search for an optimal integer solution in this fashion, and the need arises for good heuristics providing upper bounds on the optimal solution. We propose a simple heuristic based on sequential rounding.

The heuristic starts from the optimal solution $\bar{x}$ of the strengthened linear relaxation. Now, an index,

$$\{i,j\} \in \underset{\{k,l\}\in E}{\arg\min}\big\{c_{kl}(\lceil \bar{x}_{kl} \rceil - \bar{x}_{kl}) \;\big|\; \lceil \bar{x}_{kl} \rceil > \bar{x}_{kl}\big\},$$

is identified and the constraint $x_{ij} \geq \lceil \bar{x}_{ij} \rceil$ is added to the problem. The heuristic proceeds by sequentially solving the problem, checking for violated cutting planes, and rounding up variables until a feasible integer solution is obtained.

8.3.2 Problem Instances

The first problem instance is a real-life communications network provided by SONOFON, a Danish network operator. The network is a complete network on 7 nodes and hence has 21 edges. The two remaining problem instances are modified versions of real-life instances previously studied by e.g. Bienstock and Günlük [17] and Günlük [53], and used also in Chapter 7. One of these instances is a network representing the Atlanta area, containing 15 nodes and 22 edges. The other instance is a denser network representing the New York area. In this network we have 16 nodes and 49 edges, and there is no existing capacity on the edges. In the original versions of the two latter instances, two different types of facilities (i.e. low-capacity and high-capacity) are available for installation. The cost exhibited a high degree of economies to scale, though, and hence we chose to use only the high-capacity facilities for our experiments in order to fit the instances into the present context. All three problem instances have fully dense traffic matrices.

For each network we performed a series of experiments with varying number of scenarios. We considered only one type of failure, namely failure of a single edge. Moreover, we randomly generated a number of outcomes of future point-to-point demands assuming some uncertainty in the overall demand level, captured in a parameter μ, as well as some regional (node dependent) fluctuations, captured in parameters λ_i, $i \in V$. Hence for $s \in \{1,\dots,S\}$ and $k \in K^s$ demand for commodity k under scenario s was calculated as

$$D_k^s = \mu^s \lambda_{o(k)}^s \lambda_{d(k)}^s D_k,$$

where D_k is the expected demand for commodity k, $o(k)$ and $d(k)$ are the origin and destination of commodity k, respectively, and the random parameters μ^s and λ_i^s, $i \in V$, were sampled from uniform distributions,

$$\begin{aligned}&\mu^s \sim U(0.8, 1.2),\\ &\lambda_i^s \sim U(0.9, 1.1), \quad i \in V.\end{aligned}$$

For all situations with no failure, the network was required to fulfill the capacity requirements for each point-to-point pair of nodes, no matter the level of demand. Hence for a scenario $s \in \{1,\dots,S\}$ representing a situation with no failure, the binary variable was excluded from the cutting planes (i.e. set to zero) and we let $\rho_k^s = 1$ for all $k \in K^s$. Likewise, a scenario $s \in \{1,\dots,S\}$ corresponding to some failure situation was considered as violating whenever the capacity requirement was not fulfilled for some point-to-point pair of nodes, i.e. we let $\rho_k^s = 1$ for all $k \in K^s$.

Assuming that the probability of failure is equal for all edges in the network, we used uniform scenario probabilities for the failure situations, i.e. for a problem instance with $|E|$ edges in the network and d possible values of future point-to-point demands we let $\pi^s = 1/(|E| \cdot d)$ for all scenarios $s \in \{1, \ldots, S\}$ corresponding to some specific failure state. Situations with no failure were not treated as other scenarios since the corresponding binary variables were excluded from the formulation cf. the discussion above, and hence zero probability were assigned to these scenarios. Thus, we note that using these scenario probabilities, the parameter p in the p-restricted problem (8.2.2) actually denotes the *conditional* probability of capacity requirements to be violated given that a failure occurs — i.e. the fraction of failure situations for which the capacity requirements are violated.

Remark 8.3.1. The assumption that the probability of failure is equal for all edges in the network is justified in situations where typical failures mainly occur at the end-points of connections. Such failures include breakdowns of electronic equipment as well as human errors during configuration of switches. Since the backbone network is normally well-protected (e.g. carried along highways, railroads or high-voltage transmission lines), such failures are often more likely than damage to the actual connection. In particular, the assumption was found reasonable by SONOFON.

8.3.3 Computational Results

The first series of experiments were conducted in order to examine the quality of the approximations provided by the strengthened linear relaxation and the upper bounding heuristic. The first run was performed on the Atlanta problem assuming that demand is deterministic ($d = 1$). All values of p ranging from 0 to 1 were considered. Figure 8.1 shows the optimal objective values resulting from the IP-formulation, the linear relaxation, the strengthened linear relaxation and the upper bounding heuristic, respectively.

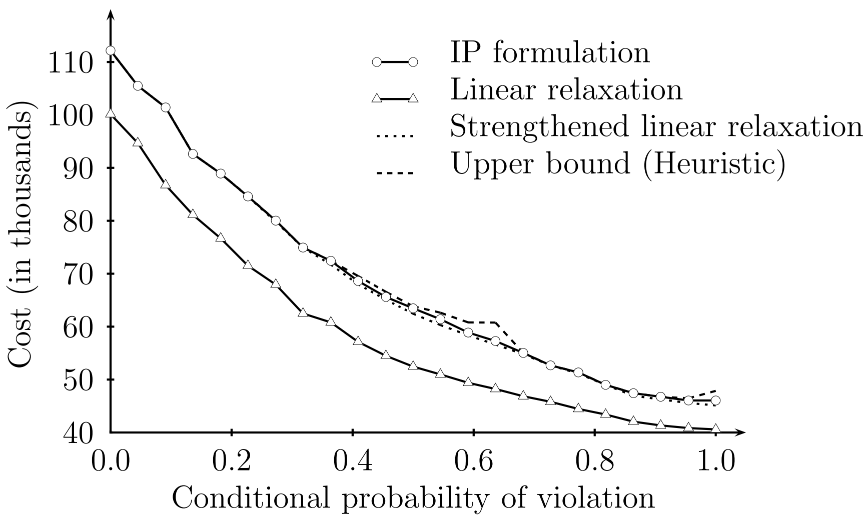

Figure 8.1: Atlanta problem ($d = 1$)

For this instance the integrality gaps were substantial, ranging from 10% to 20% for different values of p. Evidently, though, we see that the strengthened linear relaxation as well as the upper bounding heuristic performed extremely well. The integrality gap was closed by the strengthened linear relaxation for 8 of 26 problems, and the remaining gap was very small ($< 2\%$) in all cases. The upper bound was 5.7% off in the worst case, and the optimal integer solution was found by the heuristic for 14 different values of p.

The second run was performed on the SONOFON problem. Once again we considered all values of p ranging from 0 to 1, and assumed that demand is deterministic. For this instance the integrality gaps were quite small, and hence, in Figure 8.2, we plot for each value of p the optimal objective value resulting from the linear relaxation (LR), the strengthened linear relaxation (SLR), and the upper bounding heuristic (UBH), *relative* to the objective value of the optimal integer solution.

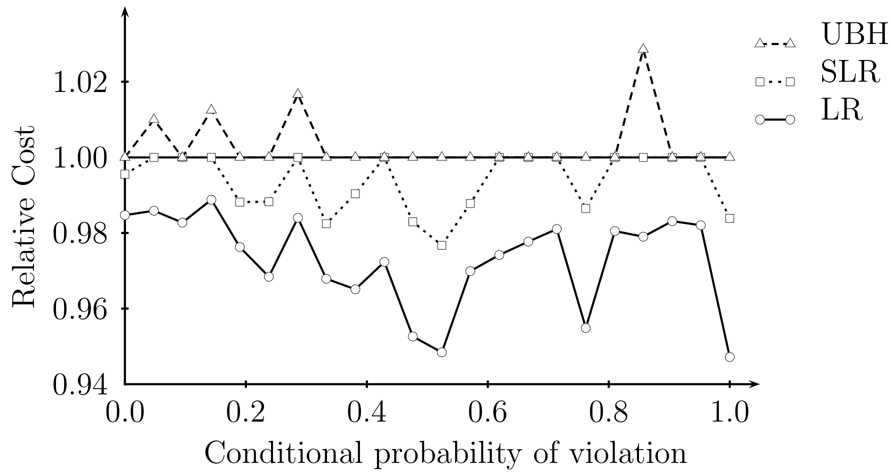

Figure 8.2: SONOFON problem ($d = 1$)

Once again we see that the strengthened linear relaxation as well as the upper bounding heuristic performed very well. In fact the integrality gap was closed by the strengthened linear relaxation for 12 of 22 problems and more than halved in most other cases, and the upper bounding heuristic found the optimal integer solution in all but four cases.

Finally, we performed one more run on the SONOFON problem, this time generating 5 possible outcomes of future demand, and considering values of p ranging from 0 to 0.25. Results are shown in Figure 8.3.

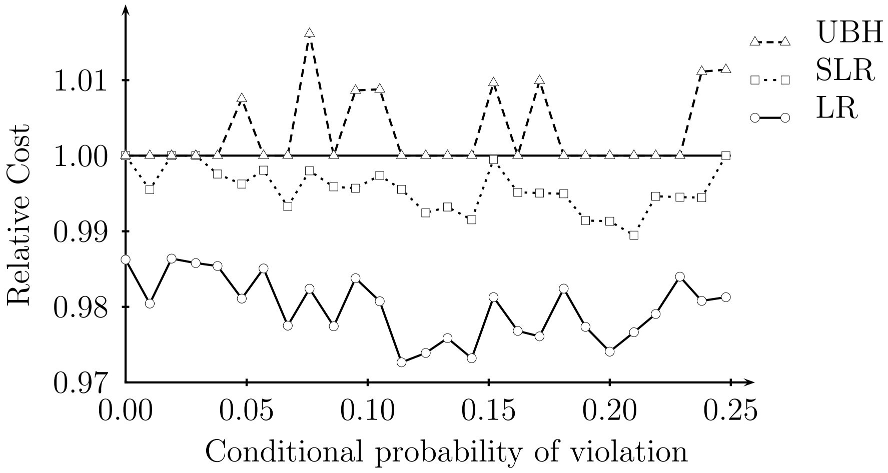

Figure 8.3: SONOFON problem ($d = 5$)

Again we see that the integrality gaps are quite small for the SONOFON problem. The integrality gap remaining from the strengthened linear relaxation was less than 1% for all different values of p and the optimal integer solution was found by the upper bounding heuristic for 19 of 27 problems.

The second series of experiments were conducted in order to test the practicability of the solution procedure. For each of the three instances we solved a series of problems with varying number of scenarios, and in all cases the maximum acceptable level of p was set to 10%. The results are presented in Tables 8.1 through 8.3. Here we report the CPU time required to solve the bicriterion stochastic programming problem (8.1.1) using two alternative versions of the solution procedure — one that only employs the strengthened linear relaxation of the master problem (8.2.4) in Algorithm 8.2, and one that solves the IP-formulation of each p-restricted problem (8.2.2), initially employing the strengthened linear relaxation of the master problem (8.2.4) and subsequently reintroducing the integer requirements on x. Also, the number of p-restricted problems to be solved during computation is reported. For illustration we also report these figures when only values of p less than or equal to 5% are considered. All CPU times are reported as minutes:seconds. Computations were stopped after three hours of CPU time and in this case the last value of p for which the p-restricted problem was being solved is given in brackets. Finally, the number of failure states is reported as $|E| \cdot d$ for a problem instance with d possible values of future point-to-point demands. We note that this number equals the number of binary variables in the master problem, whereas the number of second-stage problems to be solved is $(|E| + 1) \cdot d$, since scenarios with no failure should also be considered.

Table 8.1: SONOFON problems

Number of failure states	Maximum value of p	Number of problems	CPU time IP formulation	CPU time Relaxation
$21 \cdot 1$	0.05	2	0:01	0:01
	0.10	3	0:01	0:01
$21 \cdot 5$	0.05	6	0:11	0:04
	0.10	11	1:03	0:19
$21 \cdot 10$	0.05	11	2:35	0:58
	0.10	22	92:36	25:03

Table 8.2: Atlanta problems

Number of failure states	Maximum value of p	Number of problems	CPU time IP formulation	CPU time Relaxation
$22 \cdot 1$	0.05	2	0:07	0:06
	0.10	3	0:10	0:09
$22 \cdot 5$	0.05	6	0:53	0:41
	0.10	12	5:25	1:52
$22 \cdot 10$	0.05	12	25:33	5:44
	0.10	23	180:00 (0.077)	107:55

Table 8.3: New York problems

Number of failure states	Maximum value of p	Number of problems	CPU time IP formulation	Relaxation
$49 \cdot 1$	0.05	3	135:37	6:02
	0.10	5	180:00 (0.061)	12:25
$49 \cdot 5$	0.05	13	180:00 (0.000)	180:00 (0.037)
	0.10	25	– ‖ –	– ‖ –
$49 \cdot 10$	0.05	25	180:00 (0.000)	180:00 (0.010)
	0.10	50	– ‖ –	– ‖ –

As expected we see that computation time increases drastically with the number of scenarios as well as with the size of the network. The increase in CPU time, when a larger number of possible outcomes of future demand is generated, is partly explained by the fact that a larger number of second-stage multicommodity flow problems have to be solved in each iteration. More important, however, was the increased effort required to solve the master problems as the number of binary variables increase. When a network containing a larger number of edges is considered, on the other hand, not only the number of scenarios (and hence the number of second-stage problems and the number of binary variables in the master problem) increases, but also the number of first-stage variables. Hence the problem complexity is heavily dependent on the number of edges in the network, as illustrated by the difference in CPU time for the New York problem compared to the two smaller instances. Finally, we observed that the p-restricted problems were increasingly difficult to solve, as the value of p was increased. Hence for the New York network the algorithm was only practicable for the case of deterministic demand, unless only very small values of p were considered.

Acknowledgment

Again, the author would like to thank Oktay Günlük for providing two of the problem instances for computational testing.

Chapter 9

Deployment of Mobile Switching Centers

Mobile telecommunications network operators have been facing a rapid growth in demand for several years and this trend seems likely to continue. This development forces the network operator to constantly expand the capacity of the network in order to provide an acceptable grade-of-service to customers. There is a vast amount of literature concerning the optimal expansion of link capacities in a telecommunications network. Apart from the work in this direction presented in previous chapters of this thesis, we refer to e.g. Balakrishnan et al. [6, 7, 8] and Chang and Gavish [33, 34] for different approaches to such types of problems. The link capacities, however, do not constitute the only potential bottleneck in a telecommunications network, since capacity restrictions may be imposed not only on traffic but also on the number of customers served by the network. In this chapter we study a network design problem in which some capacity constraints are imposed to restrict traffic on links in the network while others are imposed to restrict the number of customers served by nodes in the network. As in the preceding chapters, the problem is formulated as a two-stage stochastic program in order to take due account of the inherent uncertainty involved in the assessment of future demand.

9.1 Problem Formulation

In the following sections we go through a thorough description of the network design problem that we are concerned with. First, we give a brief outline of the problem, and in particular we discuss the general structure of a mobile telecommunications network. The problem formulation is subsequently formalized, and we discuss in detail all of the parameters and variables involved before we finally present a two-stage stochastic programming formulation of the problem.

9.1.1 General Outline

The region of service is partitioned into a number of cells, each of which is served by a base transceiver station (BTS) that picks up the signal from customers' mobile phones. Each BTS is connected to one base station controller (BSC), and each BSC, on the other hand, serves a number of BTSs and is connected to one mobile switching center (MSC). Finally, each MSC serves a number of BSCs, and the MSCs are interconnected in a meshed network. A small sample network is illustrated in Figure 9.1.

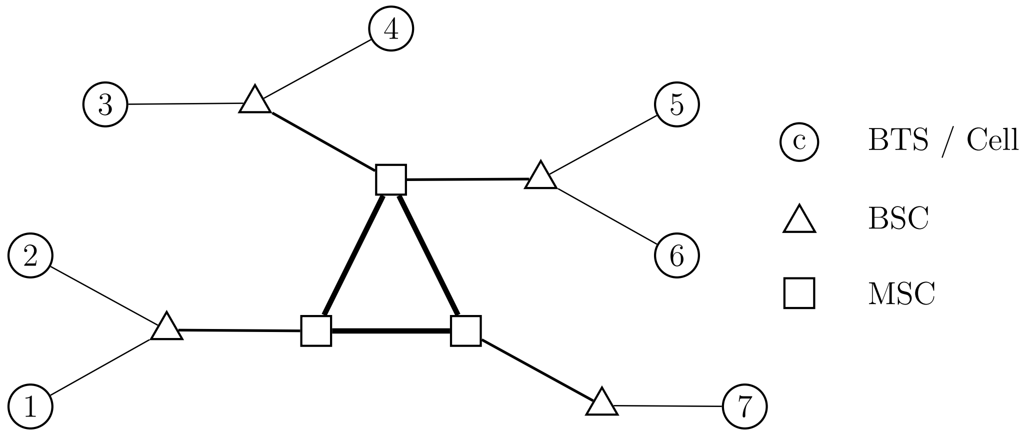

Figure 9.1: Illustration of a mobile telecommunications network.

As described above, each MSC serves a number of cells. All information about customers that are presently located in one of these cells is stored in a database in the MSC, referred to as the visitor location register (VLR). This VLR has a limited capacity, thus restricting the number of customers that can be served (through BTSs and BSCs) by an MSC. For this reason the network operator not only has to expand the link capacities but should also consider when and where to deploy new MSCs in order to be able to serve the increasing demand from customers. The BTSs and BSCs, on the other hand, are already distributed across the country so as to cover the entire region of service, and they do not constitute a potential bottleneck. Therefore we will treat the number and locations of BTSs and BSCs as exogenous input for the network design problem. Hence we consider a problem involving three major groups of decisions — deployment of a number of new MSCs, allocation of BSCs to new and existing MSCs, and capacity expansion of the transmission links interconnecting the MSCs. These decisions must be made so as to minimize the incurred costs while meeting customer demand and observing the capacity restrictions of VLRs as well as of transmission links.

The total costs incurred will consist of four terms. Associated with each group of decisions is a corresponding cost term, i.e. the cost of new MSCs, the cost of connecting BSCs to MSCs, and the cost of expanding the link capacities. In addition, we include a penalty cost for supporting so-called MSC handovers. In general, a handover occurs whenever a customer passes from one cell to another during an ongoing call. Hence, three different kinds of handovers may occur — BTS handovers occur when the two cells involved are served by the same BSC, BSC handovers occur when the two cells involved are served by two different BSCs, each of which is connected to the same MSC, and finally MSC handovers occur when the two cells involved are served by two different MSCs.

Example 9.1.1. Consider the network in Figure 9.1. Here, a BTS handover occurs for example when a customer moves from cell 1 to cell 2, a BSC handover occurs for example when a customer moves from cell 4 to cell 5, and finally, an MSC handover occurs for example when a customer moves from cell 2 to cell 3.

BTS handovers and BSC handovers are relatively easily handled, since the two cells involved are served by the same particular MSC. When an MSC handover occurs, on the other hand, the control of the ongoing call must be passed from one MSC to another, and this is a technically complicated task leading to an increased risk of loosing the call. The penalty cost is included to limit the number of such MSC handovers. Clearly, we cannot completely eliminate the occurrence of MSC handovers, but the penalty cost should be large enough to ensure that the cells served by an MSCs constitutes a geographically connected region. (The connection of BTSs to BSCs already ensures that the cells served by some particular BSC constitute a geographically connected region.)

Tzifa et al. [155] study a problem that is similar to that outlined above, but the authors consider only the access network (i.e. the tree-like network connecting BTSs to the core network), thus ignoring the capacity expansion of transmission links interconnecting the MSCs. Moreover, the problem of optimally assigning BSCs to MSCs has been addressed by several authors such as Merchant and Sengupta [91] and Saha, Mukherjee, and Bhattacharya [135]. Apart from minimizing the cost of connecting BSCs to MSCs and the handover cost, it is customary to enforce some degree of load balancing among the MSCs. Tzifa et al. and Saha, Mukherjee, and Bhattacharya explicitly include a penalty cost on uneven loads in the objective function, whereas Merchant and Sengupta propose to handle the load balancing problem parametrically. Here we will not explicitly consider load balancing, but the parametric approach of Merchant and Sengupta may easily be adopted in our setting.

The above-mentioned authors all follow a deterministic approach in the sense that the cost parameters, the number of customers, and the demand for bandwidth are assumed to be known at the point of decision. It is a fact, however, that the time that passes from the moment at which deployment of MSCs is resolved on, until the equipment is actually in place and available for use, is rather long (about a year). This means that the network operator does not have complete knowledge about several important parameters of the model at the time the decision has to be made. For this reason the definitive decision on allocation of BSCs to MSCs should be put off for as long as possible, allowing uncertainty to be at least partially revealed. This is the incentive for us to model the problem as a two-stage stochastic program with mixed-integer recourse, representing the uncertain parameters by random variables. Here the first stage consists of deployment of MSCs. At the point in time the deployment must be planned, the only available information about the uncertain parameters is assumed to be conveyed through their distribution, and hence the first-stage decision cannot be based on the actual outcome of these parameters — i.e. the decision must be *non-anticipative*. In the second stage, outcomes of all random parameters are observed and an optimal allocation of BSCs to MSCs and a corresponding routing of traffic in the resulting network are determined.

The true distribution of the random variables describing future demand and prices can at best be estimated from historical data combined with expert opinions on future development, and this distribution is most likely absolutely continuous with a multivariate distribution function. As pointed out also in the preceding chapters, though, such an approach would lead to severe computational difficulties, and hence once again we employ a scenario approach, representing the uncertain outcome of random parameters by a finite number of scenarios with prescribed probabilities of occurrence.

9.1.2 Parameters

We will consider a finite number of potential locations for new MSCs and hence the nodes of the network will be described by three finite sets representing the locations of MSCs and BSCs.

- V_1 : The set of locations of existing MSCs.
- V_2 : The set of potential locations for new MSCs.
- W : The set of locations of BSCs.

Note that a given location may very well be represented as a node in more than one of the sets (even in all of them). In fact, the model allows for a single location to be represented as several nodes in one set, for example if we consider deploying more than one MSC at a location. Now, the network interconnecting the MSCs is modeled as a connected undirected graph $G = (V, E)$, where the node set $V = V_1 \cup V_2$ represents the existing and potential locations of MSCs, and the edge set E represents the existing and potential links $\{i, j\}$ between nodes $i, j \in V$.

There are two groups of demand input for the model — demand for bandwidth on links of the network, and demand for VLR-capacity at the nodes of the network. As pointed out also in the preceding chapters, modeling the actual process of real-time call-by-call routing within a long-term planning model, as the one considered here, is obviously not viable. Hence, the demand for bandwidth is not considered as individual telephone calls, but is given as a set of capacity requirements between node-pairs, needed to maintain a prescribed grade-of-service for customers.

Remark 9.1.1. As discussed in Remark 7.2.1 on page 96, a possible way of transforming a general stochastic process of demands into a static capacity requirement was proposed by Dempster, Medova, and Thompson [40] and Medova [90]. Here the authors consider ATM-based broadband integrated services digital networks (B-ISDN) and use a chance-constrained stochastic programming approach to determine the capacity requirements for each point-to-point pair as the effective bandwidth requirements needed to ensure that a set of blocking probabilities are not exceeded.

We will consider demand for bandwidth at BSC level and we assume that all traffic is bidirectional. For modeling purposes, however, it is much more convenient to work with directed flow. Therefore we assign an arbitrary direction to each point-to-point demand and refer to its origin and destination. In particular, given some numbering of the BSCs, we assume that all traffic between some particular pair of BSCs originates at the lower numbered BSC and terminates at the higher numbered BSC. In other words, rather than saying that a certain amount of bandwidth is required for calls between some particular pair of BSCs, we will say that an equivalent amount of flow should be sent from the lower numbered BSC to the higher numbered BSC. Still, to allow for the appropriate routing of bidirectional traffic, edge capacities are dimensioned with respect to the total traffic on a given edge, disregarding the arbitrarily assigned directions of flow. As in the preceding chapters, the demand for bandwidth on the links of the network is described by a set K of commodities, and hence the demand input for the model is given by the following sets of parameters.

- D^s_{kr} : The net demand for commodity k at BSC r under scenario s ($k \in K$, $r \in W$, $s \in \{1, \ldots, S\}$).
- L^s_r : The load of BSC r under scenario s on the VLR in the MSC to which it is connected ($r \in W$, $s \in \{1, \ldots, S\}$).

Remark 9.1.2. Two main approaches for defining the set of commodities K have been used in the literature. A disaggregated formulation defines a commodity for each point-to-point demand, resulting in a total of $O(|W|^2)$ commodities. Such an approach was followed in Chapter 8. In the present setting we generally find it more convenient, though, to reduce the number of variables by working with an aggregated formulation containing a total of only $O(|W|)$ commodities. This is achieved by letting each commodity correspond to demand originating at a given BSC with respect to the arbitrary directions assigned to traffic. A similar approach was followed in Chapter 7 in connection with the capacitated network design problem. One should note, however, that if survivability constraints are to be imposed, the more detailed description of traffic provided by the disaggregated formulation may prove favorable.

Remark 9.1.3. Consider some particular scenario $s \in \{1, \ldots, S\}$. We emphasize that for all $k \in K$ and $r \in W$ the parameter D^s_{kr} is the *net demand* for commodity k at BSC r under scenario s, and hence $\sum_{r \in W} D^s_{kr} = 0$ for all $k \in K$. Furthermore, for $k \in K$ and $r \in W$ we have $D^s_{kr} < 0$ if and only if BSC r is the origin of commodity k, and in that case the parameter represents the total demand for commodity k. Using the aggregated formulation of commodities, we have for all $k \in K$ and $r \in W$ such that BSC r is *not* the origin of commodity k, that the parameter D^s_{kr} is non-negative and represents the amount of bandwidth that is required for calls between the origin of commodity k and BSC r. Employing the disaggregated formulation, on the other hand, we have for $k \in K$ and $r \in W$ that $D^s_{kr} > 0$ if and only if BSC r is the destination of commodity k.

Corresponding to the two types of demand, we have two types of existing capacity in the network — capacity restricting flow on edges of the network and capacity restricting the number of customers served by nodes in the network.

- C_{ij} : Current flow-capacity of the edge $\{i, j\}$ ($\{i, j\} \in E$).
- M_i : VLR-capacity of the MSC located at node i ($i \in V$).

The cost structure is described by the following sets of parameters some of which are treated as deterministic, while others are assumed to be uncertain at the point in time the decision has to be made, thus depending on the future scenario that will occur.

- c_i : The cost of deploying an MSC at node i ($i \in V_2$).
- p^s_{ij} : The cost of adding one unit of capacity on the edge $\{i, j\}$ under scenario s ($\{i, j\} \in E$, $s \in \{1, \ldots, S\}$).
- q^s_{ri} : The cost of connecting BSC r to node i under scenario s ($r \in W$, $i \in V$, $s \in \{1, \ldots, S\}$).
- h^s_{rt} : The penalty cost (for supporting handovers) incurred if BSC r and BSC t are connected to different MSCs under scenario s ($r, t \in W$, $r < t$, $s \in \{1, \ldots, S\}$).

Remark 9.1.4. Note that we assume the cost of expanding the capacity of a connection to be linear and that we do not include a fixed cost for establishing the connection. The reason for this is the fact that SONOFON, the company in cooperation with which this research project was engaged upon, had already available a physical network with sufficient link capacities. In order to utilize this capacity, however, it may be necessary to install additional equipment at the end-points of connections, and this cost is assumed to be linear with respect to the capacity provided. It would certainly be more appropriate to assume that capacity could only be installed in multiples of a fixed batch size, cf. Chapter 8 where another capacity expansion model was considered for the same particular network. Since the capacity expansion of transmission links is only of secondary importance in the present context, though, we found the approximation provided by the linear cost function satisfactory.

Remark 9.1.5. Regarding the penalty cost for supporting handovers, it should be noted that for all $r, t \in W$ with $r < t$ and for $s \in \{1, \ldots, S\}$, the parameter h^s_{rt} should reflect the frequency of customers crossing from the region served by BSC r to the region served by BSC t or vice versa during an ongoing call.

9.1.3 Variables

The main decisions to be made are deployment of new MSCs and allocation of BSCs to MSCs. These decisions are represented by the following two sets of binary variables.

$$\cdot\ x_i = \begin{cases} 1 & \text{if an MSC is deployed in node } i; \\ 0 & \text{otherwise,} \end{cases} \qquad i \in V_2.$$

$$\cdot\ y^s_{ri} = \begin{cases} 1 & \text{if BSC } r \text{ is connected to node } i; \\ 0 & \text{otherwise,} \end{cases} \qquad r \in W,\ i \in V,\ s \in \{1, \ldots, S\}.$$

As indicated by the dependency of the variables y on the future scenario, the allocation of BSCs to MSCs is allowed to depend on the outcome of the random parameters. That is, the decision on allocation of BSCs to MSCs is postponed to the second stage to take full advantage of the additional information which is available at this point.

Finally, the following sets of variables are used to describe flow in the network, and the capacity expansion of links needed to carry this flow. Since flow does not occur until demand is realized, these variables all belong in the second stage.

- f^s_{ijk} : Flow of commodity k on edge $\{i, j\}$ in direction from i to j under scenario s ($k \in K$, $\{i, j\} \in E$, $s \in \{1, \ldots, S\}$).
- f^s_{jik} : Flow of commodity k on edge $\{i, j\}$ in direction from j to i under scenario s ($k \in K$, $\{i, j\} \in E$, $s \in \{1, \ldots, S\}$).
- v^s_{ij} : Additional capacity to be installed on edge $\{i, j\}$ under scenario s ($\{i, j\} \in E$, $s \in \{1, \ldots, S\}$).

Remark 9.1.6. The variables f are described as directed flow but, as previously discussed, their practical interpretation is somewhat different. Thus, recall that the traffic flow was introduced to represent the bandwidth requirements between pairs of BSCs, in the sense that an amount of flow equivalent to the required bandwidth is routed between each pair of BSCs. Hence a flow of some commodity $k \in K$ in any direction on an edge of the network, simply implies that an equivalent amount of bandwidth on that edge should be designated to the demand for commodity k. Also, note that it is easy to extract an individual routing of the point-to-point demands by disaggregating f.

9.1.4 Two-Stage Formulation

We are now ready to formulate the problem of optimally deploying a number of new MSCs and allocating BSCs to MSCs as a two-stage stochastic program. For $s \in \{1, \dots, S\}$ we assume that the probability of scenario s to actually occur is known, and we denote it by π^s. The first-stage objective is to minimize the sum of the cost of new MSCs and the expected value of the cost incurred in the second stage,

$$\min \sum_{i \in V_2} c_i x_i + \sum_{s=1}^{S} \pi^s Q^s(x) \tag{9.1.1a}$$

$$\text{s.t. } x \in \mathbb{B}^{|V_2|}. \tag{9.1.1b}$$

Here, for $s \in \{1, \dots, S\}$, the second-stage value function Q^s is given by

$$Q^s(x) = \min \sum_{r \in W} \sum_{i \in V} q^s_{ri} y^s_{ri} + \sum_{\{i,j\} \in E} p^s_{ij} v^s_{ij} + \sum_{\substack{r,t \in W \\ r<t}} h^s_{rt} \sum_{i \in V} (y^s_{ri} - y^s_{ti})^+ \tag{9.1.2a}$$

$$\text{s.t. } \sum_{r \in W} L^s_r y^s_{ri} \le M_i, \qquad i \in V_1, \tag{9.1.2b}$$

$$\sum_{r \in W} L^s_r y^s_{ri} \le M_i x_i, \qquad i \in V_2, \tag{9.1.2c}$$

$$\sum_{i \in V} y^s_{ri} = 1, \qquad r \in W, \tag{9.1.2d}$$

$$\sum_{j:\{i,j\} \in E} f^s_{jik} - \sum_{j:\{i,j\} \in E} f^s_{ijk} = \sum_{r \in W} D^s_{kr} y^s_{ri}, \qquad i \in V,\ k \in K, \tag{9.1.2e}$$

$$\sum_{k \in K} \left(f^s_{ijk} + f^s_{jik} \right) \le C_{ij} + v^s_{ij}, \qquad \{i,j\} \in E, \tag{9.1.2f}$$

$$y^s \in \mathbb{B}^{|W||V|},\ v^s \in \mathbb{R}^{|E|}_+,\ f^s \in \mathbb{R}^{2|E||K|}_+. \tag{9.1.2g}$$

We have used the notation a^+ to denote $\max\{0, a\}$ for $a \in \mathbb{R}$, and hence for $r, t \in W$ and $s \in \{1, \dots, S\}$ the third term of the second-stage objective (9.1.2a) includes the handover cost between BSCs r and t if and only if these BSCs are allocated to different MSCs under scenario s. The constraints (9.1.2b) and (9.1.2c) ensure that the total load from the BSCs connected to an MSC does not exceed the capacity of the VLR. Moreover, (9.1.2c) ensures that a BSC can only be connected to an MSC if this is actually deployed ($x_i = 1$), while (9.1.2d) ensures that all BSCs are connected to exactly one MSC. Next, (9.1.2e) is a flow conservation constraint stating that the net flow of a commodity into some MSC should equal the aggregate net demand for the commodity from BSCs connected to that MSC.

Finally, the constraint (9.1.2f) states that the aggregate flow on an edge cannot exceed the total capacity installed on the edge.

Remark 9.1.7. We note that the nonlinear term in the second-stage objective may easily be replaced by a linear one. Hence for all $r, t \in W$ with $r < t$ and for $s \in \{1, \ldots, S\}$, we let H^s_{rt} be a variable representing the handover cost incurred between BSCs r and t under scenario s. These variables may be appropriately defined within the model using a number of linear constraints,

$$H^s_{rt} \geq h^s_{rt}(y^s_{ri} - y^s_{ti}) \qquad i \in V,\ r, t \in W,\ r < t,\ s \in \{1, \ldots, S\}, \tag{9.1.3}$$

and the nonlinear term may be replaced by a simple summation of the new variables. Thus, if the constraints (9.1.3) are added in the definition of $Q^s(x)$ for $s \in \{1, \ldots, S\}$, then the third objective term may be replaced by

$$\sum_{\substack{r,t \in W \\ r<t}} H^s_{rt}.$$

9.2 Solution Procedure

As already pointed out, problem (9.1.1)-(9.1.2) is a two-stage stochastic program with binary first stage and mixed-integer recourse, and hence in principle it may be solved as such by any of the general purpose algorithms for such problems discussed in Section 3.3. In the application of the model to the network of SONOFON, however, we chose simply to enumerate the first-stage solutions. This was done for the following reasons. First, the first-stage decision x is a binary vector of relatively small dimension, and hence the number of possible solutions is limited. Second, the cost of a new MSC is orders of magnitude higher than any other cost parameter, implying that an optimal deployment of MSCs consists of a minimum number of new MSCs providing enough capacity to satisfy demand. Since only one or two new MSCs were required to provide sufficient capacity in the instances we considered, this means that the number of solutions to be considered is very limited. Finally, when the first-stage solutions are simply enumerated, all that is needed is a number of evaluations of the functions Q^s, $s = 1, \ldots, S$. These evaluations are relatively easily performed for fixed values of x, in particular when only one or two new MSCs are to be deployed, since in that case all variables and constraints in problem (9.1.2) that are related to MSCs which are not to be deployed, may simply be eliminated from the formulation.

Algorithm 9.1

Step 1 (*Initialization*) Let $n = 0$ and $\bar{z} = \infty$.

Step 2 (*Enumeration*) Let the candidate list $\mathcal{L}$ consist of the $\binom{|V_2|}{n}$ first-stage solutions corresponding to the deployment of exactly n new MSCs.

Step 3 (*Evaluation*) Let $\bar{z} = \min\limits_{x \in \mathcal{L}} \Big\{ \sum_{i \in V_2} c_i x_i + \sum_{s=1}^{S} \pi^s Q^s(x) \Big\}$.

Step 4 (*Termination*) If $\bar{z} < \infty$, stop; the solution that yielded the upper bound $\bar{z}$ is optimal. Otherwise, let $n = n + 1$ and return to Step 2.

Remark 9.2.1. Clearly, enumeration may not always be the preferred alternative for solving problem (9.1.1)-(9.1.2). If, for example, one considers a *greenfield case* where no MSCs are currently in place, an enumeration tree of considerable size may have to be created, and hence it may prove advantageous to resort to one of the general purpose algorithms discussed in Section 3.3. In particular, the dual decomposition procedure, outlined on page 42, may be directly applied to solve the problem. With binary first-stage decisions this algorithm is guaranteed to terminate with an optimal solution in a finite number of iterations. Moreover, the procedure involves the solution of scenario subproblems with $|V_2| + n_2$ variables and m_2 constraints, where n_2 and m_2 denotes the number of variables and constraints, respectively, in the second-stage problem (9.1.2). Hence, because of the small dimension of the first-stage solution vector, the size of these scenario subproblems is not critically increased compared to the size of the second-stage problems.

In order to solve problem (9.1.1) using enumeration of the first-stage solutions as suggested above, we need an efficient procedure for solving the second-stage problems (9.1.2). To this end we applied the concept of branch-and-cut which has proved to be a powerful tool for the solution of (mixed-) integer programming problems. As in ordinary branch-and-bound, we start with the LP-relaxation of the mixed-integer programming problem and build a partitioning of the solution space in order to obtain an integral solution. The crucial idea in branch-and-cut is to combine this approach with a continuous generation of cutting planes tightening the formulation and thus reducing the size of the branching tree. For a thorough discussion of the branch-and-cut approach, we refer to Padberg and Rinaldi [105] and Günlük [53].

9.2.1 Valid Inequalities

In this section we consider two classes of valid inequalities that proved useful for the solution of problem (9.1.1)-(9.1.2). First of all we consider an inequality based on the total VLR-capacity installed through deployment of new MSCs. The inequality simply states that the total capacity of all VLRs in the resulting network should exceed the total demand from all BSCs. Formally the inequality is derived by summing the constraints (9.1.2b)-(9.1.2c), rearranging, and rounding.

Proposition 9.2.1. *Let* $M = \max_{i \in V_2}\{M_i\}$. *The following inequality is valid for the feasible region* $K_2 = \left\{x \in \mathbb{B}^{|V_2|} \mid \sum_{s=1}^{S} \pi^s Q^s(x) < \infty\right\}$,

$$\sum_{i \in V_2} x_i \geq \max_{s \in \{1,\dots,S\}} \left\lceil \frac{1}{M}\Big(\sum_{r \in W} L_r^s - \sum_{i \in V_1} M_i\Big) \right\rceil.$$

Remark 9.2.2. As mentioned in the previous section, we used an enumeration scheme to solve the problem instance considered in Section 9.3. Hence the above inequality was not actually included in the formulation but was merely used at initialization of Algorithm 9.1, letting $n = \max_{s \in \{1,\dots,S\}} \left\lceil \frac{1}{M}\big(\sum_{r \in W} L_r^s - \sum_{i \in V_1} M_i\big)\right\rceil$. If, however, one was to employ the dual decomposition procedure as briefly discussed in Remark 9.2.1, the inequality should be included in all scenario subproblems.

As cutting planes in the branch-and-cut procedure used to solve the second-stage problems, we used the following class of valid inequalities. The inequalities are based on the VLR-capacity of the individual MSCs and can be used to enforce the fact that each BSC must be allocated to a unique MSC. Again, the underlying idea of the inequalities is simple — if the total demand from a group of BSCs exceeds the VLR-capacity of an MSC, we cannot allocate all of these BSCs to the MSC in question. More formally, the constraints (9.1.2b) and (9.1.2c) give rise to standard cover inequalities derived for knapsack constraints (see e.g. Nemhauser and Wolsey [97]).

Proposition 9.2.2. *Let U be a subset of W such that $\sum_{r \in U} L_r^s > M_i$ for some MSC $i \in V$ and some scenario $s \in \{1, \ldots, S\}$. Then the following inequality is valid for the feasible region of the second-stage problem (9.1.2),*

$$\sum_{r \in U} y_{ri}^s \leq |U| - 1.$$

Remark 9.2.3. Naturally, this inequality will only be useful when the subset U of W is minimal in the sense that $\sum_{r \in U \setminus \{t\}} L_r^s \leq M_i$ for all $t \in U$, since it is otherwise dominated by other inequalities of the same type.

9.3 Computational Experiments

In this section we describe the practical application of our model. We have implemented the solution procedure in C++ using procedures from the callable library of CPLEX 6.6 to solve the second-stage problems. The algorithm was used to solve a real-life problem provided by SONOFON.

9.3.1 Problem Instance

The network of SONOFON has 7 existing MSCs, 33 BSCs, and 14 potential locations for new MSCs. The network interconnecting the existing MSCs is complete and we do not preclude any potential edges from consideration when new MSCs are installed — hence the edge set E consists of 210 edges. In the implementation, the number of binary variables was reduced by dividing the area of interest into a number of regions and precluding from consideration certain allocations of BSCs to MSCs across regions. In the resulting formulation, each second-stage problem has 433 binary variables, 14598 continuous variables, and 12045 constraints.

The cost of connecting a BSC to an MSC was set to zero if the BSC is currently connected to this particular MSC, and otherwise the total cost of a movement was estimated. Furthermore, as previously discussed, the cost of expanding link capacities is given by the total cost of installing new equipment at the end-points of connections. The issue of determining an appropriate level for the artificial penalty cost for handovers, however, is a more complicated matter. Setting this level too low, may result in a solution with a large number of handovers, which is not acceptable from a practical viewpoint. A high level, on the other hand, may result in configurations for which the gained practicability obtained by reducing the number of handovers is not sufficient to justify the increased installation cost. In practice we chose to adjust the handover costs, observing their effect on solutions, so as to ensure that the BSCs allocated to each particular MSC constitute a geographically connected region.

The current demand for bandwidth and VLR-capacity was estimated from observations of traffic and the number of customers, respectively. As previously explained, the demand for bandwidth is given as a capacity requirement between each pair of BSCs. The aggregate commodities were defined by introducing a numbering of the BSCs and treating the capacity requirement between a given pair of BSCs as a demand for traffic to be routed from the lower numbered BSC to the higher numbered BSC. Hence the demand matrix $\{D_{kr}\}_{k\in W, r\in W}$ is upper triangular, and for all $k, r \in W$ with $k < r$ the demand for commodity k at BSC r is in fact the capacity requirement between BSCs k and r. Now, future demand was calculated using the estimates of current demand scaled by different scenario dependent growth factors. We have used the following expression to generate demand for VLR-capacity from each BSC under each scenario,

$$L_r^s = \mu^s \cdot \rho_r^s \cdot L_r, \quad r \in W,\ s \in \{1, \dots, S\}.$$

Here, for $r \in W$ the parameter L_r denotes the current demand for VLR-capacity from BSC r. Moreover, for $s \in \{1, \dots, S\}$ the parameter μ^s is used to reflect the average growth in demand for VLR-capacity for the entire region of service, while for $r \in W$ and $s \in \{1, \dots, S\}$ the parameter ρ_r^s is used to reflect regional fluctuations from this average growth. To capture the correlation between the demand for VLR-capacity and the demand for bandwidth, we calculated the net demand for commodity k at BSC r under scenario s for $k \in K$, $r \in W \setminus \{k\}$, and $s \in \{1, \dots, S\}$, using current demand D_{kr}, the above-mentioned parameters reflecting growth in the number of customers, and a third parameter σ^s reflecting growth in the demand for bandwidth per customer,

$$D_{kr}^s = \mu^s \cdot \sqrt{\rho_k^s \cdot \rho_r^s} \cdot \sigma^s \cdot D_{kr}, \quad k \in K,\ r \in W \setminus \{k\},\ s \in \{1, \dots, S\}.$$

Finally, the total demand for each commodity was calculated as

$$D_{kk}^s = -\sum_{r \in W\setminus\{k\}} D_{kr}^s, \quad k \in K,\ s \in \{1, \dots, S\}.$$

Likewise, the different cost terms were generated for each scenario by introducing stochastic fluctuations on future prices. The growth factors were all sampled from uniform distributions reflecting the expectations of SONOFON for the time horizon under consideration. Since scenarios were generated by sampling, we used uniform scenario probabilities, i.e. $\pi^s = 1/S$ for $s \in \{1, \dots, S\}$.

Remark 9.3.1. As previously pointed out, the second-stage decision of allocation of BSCs to MSCs is to be made after one year, and this was the time horizon used when estimating growth factors for the cost terms. As for customer demand, however, we have used a longer time horizon when estimating the appropriate growth factors. This was done to ensure a somewhat stable solution guaranteeing sufficient network capacity for a few additional years beyond the completed deployment of new MSCs. This means, however, that demand is in fact only partially revealed at the time the second-stage decisions are to be made. Still, since the additional information available at this point will provide an improved estimate of the true rate of growth in demand, the gain of postponing some decisions to the second stage is likely to be considerable.

9.3.2 Computational Results

A series of computational experiments were performed on the network of SONOFON. In accordance with the precepts described above, we randomly generated instances with 5, 10, 50, and 100 scenarios, respectively, using a four-year time horizon for the estimation of growth factors for demand. We also considered the expected value problem (EVP), replacing all random parameters by their expected values. To investigate the effect of the valid inequalities discussed in Section 9.2.1, each of the instances was solved using two different versions of the algorithm — one that employs the cover inequalities of Proposition 9.2.2 for the solution of the second-stage problems, and one that does not. (We should mention that for these instances the valid inequality of Proposition 9.2.1 implied the deployment of just one new MSC, and hence the effect of the inequality in reducing the size of the enumeration tree was negligible.) Results of the experiments are reported in Table 9.1. For each instance we report the total number of second-stage problems solved during computation. We note that some of the first-stage solutions in the enumeration tree lead to infeasible second-stage problems, and non of these problems are counted as subproblems solved. We also report the average number of cover inequalities added to each second-stage problem and the average number of branching nodes required to solve these problems. Finally, the total CPU time spent by the procedure is reported as minutes:seconds.

Table 9.1: Computational Results (four-year time horizon)

S	Subproblems solved	Average number of cover cuts	Average number of branching nodes[a]	Overall CPU time[a]
(EVP) 1	8	22	308 (424)	0:38 (0:41)
5	40	20	307 (465)	2:35 (3:08)
10	77	19	264 (431)	4:38 (5:51)
50	367	20	336 (631)	27:38 (37:13)
100	732	20	329 (564)	52:55 (72:34)

[a]Results when no cover cuts are employed are given in brackets.

As can be seen from Table 9.1, the cover inequalities of Proposition 9.2.2 effectively helped to reduce the average number of branching nodes and hence the overall CPU time. The optimal solution of the instances with 5, 10, 50, and 100 scenarios were identical, all suggesting the deployment of one new MSC at the same particular location. The optimal solution of the expected value problem, on the other hand, also suggested the deployment of one new MSC but at a different location. Evaluating the solution of the expected value problem in the objective of the stochastic program with 100 scenarios, it turns out that we achieve a 3.5% reduction in the expected number of MSC handovers and a 2.1% reduction in the expected capacity expansion cost by solving the stochastic program rather than the expected value problem. Thus, the computational experiments clearly illustrate the importance of imitating the dynamics of the decision process through the stochastic programming model, postponing the second-stage decisions until uncertainty has been revealed rather than basing all decisions simply on the expected value of random parameters.

We also generated a number of instances using a six-year time horizon for the estimation of growth factors for demand. In this case, the valid inequality of Proposition 9.2.1 implied the deployment of at least two new MSCs for all instances with 5, 10, 50, or 100 scenarios, respectively, and hence for these instances all nodes corresponding to the deployment of just one new MSC could be discarded from the enumeration tree at initialization of the algorithm. Computational results are reported in Table 9.2.

Table 9.2: Computational Results (six-year time horizon)

S	Subproblems solved	Average number of cover cuts	Average number of branching nodes[a]	Overall CPU time[a]
1 (EVP)	1	60	3035 (3389)	0:36 (0:35)
5	128	32	644 (1060)	23:06 (28:31)
10	183	30	677 (1090)	35:53 (44:54)
50	618	39	1130 (1809)	200:54 (240:46)
100	981	41	1243 (1948)	353:44 (421:56)

[a]Results when no cover cuts are employed are given in brackets.

Using the six-year time horizon, the importance of using a stochastic programming model becomes even more obvious since in this case the solution of the expected value problem suggested the deployment of just one new MSC and hence turned out to be infeasible for all of the instances of the stochastic program with 5, 10, 50, and 100 scenarios. The optimal solution of these instances, on the other hand, all suggested the deployment of two new MSCs at the same particular locations.

Chapter 10

Internet Protocol Network Design

The foundation of IP (internet protocol) was laid in the late 1960s as the US Department of Defense sought to create a network resilient enough to withstand an enemy attack. The ARPANET (Advanced Research Projects Agency Network), initially connecting four US universities, has since then grown to what is known today as the internet. The rapid growth of the internet and its use alone provides a constantly increasing source of traffic to be carried over the IP networks of today, and moreover, IP is expected to serve as a general platform for providing data and telecommunications services in the future. Hence the problem of constructing IP networks, providing sufficient capacity for the rising demand, in a cost-efficient way, is of great importance for network providers. For a brief introduction to the concepts and terms related to IP networks, we refer to Challinor [32]. In this chapter we consider the IP network of TDC, the largest Danish network operator. Due to historical reasons, the number of IP POPs (points of presence) in the network has reached a level believed to be too high. To point out potential IP POPs for dismantling, we consider a network planning problem concerning dimensioning of the IP POPs and capacity expansion of the transmission links of the network. This problem is formulated as a two-stage stochastic program with linear recourse, using a finite number of scenarios to describe the uncertain outcome of future demand.

10.1 Problem Formulation

In the following sections we go through a thorough description of the IP network design problem. First we give a brief outline of the problem before formalizing the problem formulation and discussing in detail all of the parameters and variables involved. Finally, we present a two-stage stochastic programming formulation of the problem.

10.1.1 General Outline

The IP network basically consists of a large number of IP POPs interconnected by a number of transmission links in the form of optical fiber cables with SDH (synchronous digital hierarchy) equipment. With IP, data to be transmitted to some destination in the network is broken down into a number of small datagrams or packets, each of which is addressed with the destination before being passed into the network. The packets are sent from one IP POP to another through the network, with each IP POP examining the destination address to decide where to send the packet next. Hence, the IP POPs serve two main purposes — they handle the routing of traffic in the network and they serve as access points to the network for customers. During the period of time when the network was built up, forming its present structure, customers accessed the network by modem through the PSTN (public switched telephone network) and access was charged as regular telephone calls. Since at that time, telephone calls in Denmark where classified as either short-distance or long-distance and charged accordingly, it was felt by TDC that all customers should be able to access the internet at the lower short-distance rate. This policy has resulted in a network with a large number of IP POPs (approximately 200) distributed across the country. Today, however, a variety of internet products is offered to customers, providing alternative technologies for access as well as several different charging schemes, all of which are independent of the physical location of the IP POP providing access for the customer to the network. Moreover, all IP POPs in the network must be maintained and, more importantly, upgraded so that sufficient capacity to switch the increasing volume of IP traffic is available. Since the total amount of switching in the network (given a certain amount of traffic) increases with the number of IP POPs in the network, these considerations have lead TDC to believe that it may be economically and practically profitable to dismantle some of the IP POPs in outer, sparsely populated regions. To point out potential IP POPs for dismantling, we formulate the network design problem of TDC as a mathematical programming problem, taking into account the maintenance and upgrading of IP POPs, the connection of customers to the network, and the capacity expansion of transmission links.

To plan the design of the IP network, it is essential to have a qualified estimate of the future number of customers as well as the future volume of IP traffic to be carried over the network. Bearing in mind the rapid growth of the internet and its use, and the fact that new services to be carried over IP networks frequently emerge, it is clear that such an estimate is not readily available. In other words, the assessment of future demand inevitably involves a large degree of uncertainty that should be taken into account in the formulation of the problem, so that the performance of the resulting network is not too sensitive with respect to the actual outcome of future demand. Therefore, we employ a stochastic programming approach, treating the future number of customers and the future volume of IP traffic as random variables. Now, the network design problem fits into the class of two-stage stochastic programming problems with linear recourse, where the decisions are divided into two groups — a group of first-stage decisions that must be taken without certain knowledge about the outcome of random parameters, and a group of second-stage decisions that may be postponed until the actual outcome of random demand has been observed. In the present context, the first stage corresponds to the decisions on network design that must be planned some time ahead and hence have to be based solely on the uncertain estimates of future demand, whereas the second stage corresponds to the routing of IP traffic in the resulting network which is naturally postponed until demand has actually occurred.

Typically, it is most natural to think of the probability distribution of future demand as absolutely continuous. To handle the problem computationally, however, we employ a scenario approach as in the preceding chapters, replacing this absolutely continuous distribution with a discrete one. Hence, the uncertain outcome of future demand is described by a finite number of scenarios with prescribed probabilities of occurrence.

Remark 10.1.1. In fact, as described in Section 10.3.1, scenarios were generated for the IP network of TDC, not in an attempt to approximate some true probability distribution of random demand, but merely to represent the spectrum of possibly future outcomes of demand. Hence, some may say that our stochastic programming model should rather be referred to as a *robust optimization* problem (see Mulvey, Vanderbei, and Zenios [96]), cf. Remark 7.2.2 on page 97.

10.1.2 Network Representation

We start off with a conceptual description of the current network, facilitating the formulation of the network design problem as a mathematical program. First of all the region serviced by the network is partitioned into a number of subregions corresponding to the service areas of current IP POPs, so that all customers in any subregion are currently connected to the network through the same particular IP POP. Next, we will distinguish between two different network segments — the core network and the distributed network. The core network is a meshed network interconnecting a number of large IP POPs using SDH STMs (synchronous transfer modules). The transmission rates are STM-1, running approximately 155 Mbit/s (equivalent to an OC3), STM-4, running approximately 622 Mbit/s (equivalent to an OC12), and STM-16, running approximately 2.5 Gbit/s (equivalent to an OC48). The distributed network, on the other hand, consists of a large number of smaller IP POPs, each of which is connected to the rest of the network by either ATM (asynchronous transfer mode) PVCs (permanent virtual circuits) or a number of E1 (2 Mbit/s) circuits. For now, we will assume that each IP POP in the distributed network is connected to the rest of the network by two alternatively conveyed links of equal type and capacity. (In reality things are in fact a bit more complicated as discussed in Section 10.3.1.) A small sample IP network is illustrated in Figure 10.1.

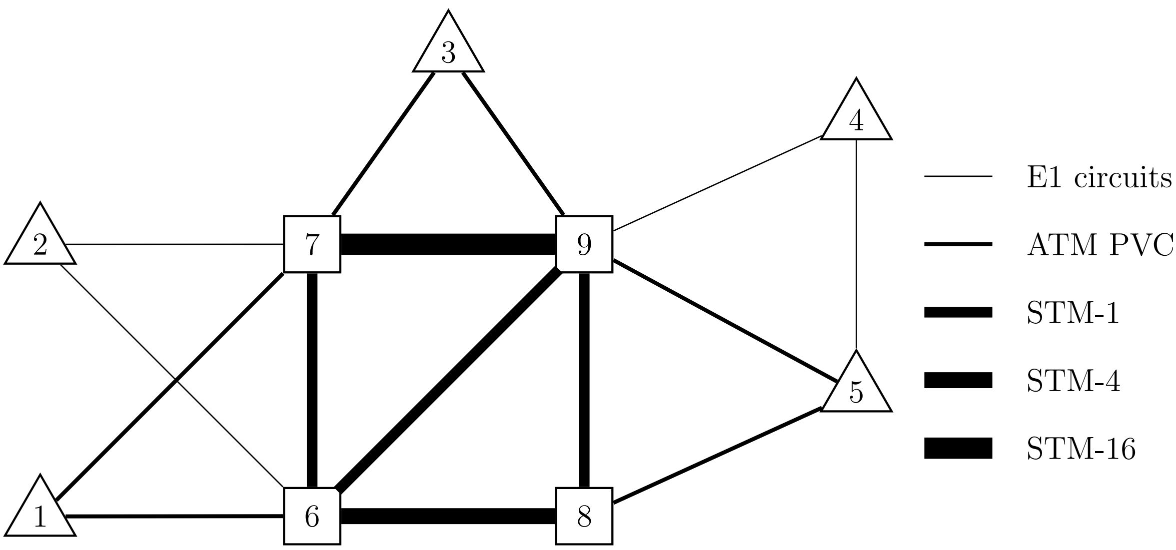

Figure 10.1: Illustration of a small IP network.

The network will be represented by a connected undirected graph $G = (V, E)$. The node set V represents the set of regions corresponding to current IP POPs, and hence a node $i \in V$ corresponds to a region in which all customers are currently connected to some particular IP POP. The edge set E, on the other hand, represents the set of transmission links in the network, and hence each edge $\{i, j\} \in E$ corresponds to a transmission link between IP POPs in regions $i \in V$ and $j \in V$. The partition of the network into the core network and the distributed network is given by the following sets.

- V_1 : The set of regions corresponding to IP POPs in the distributed network.
- V_2 : The set of regions corresponding to IP POPs in the core network.
- $E_1(i)$: The set of transmission links connecting the IP POP in region i to the rest of the network ($i \in V_1$).
- E_2 : The set of internal transmission links in the core network.

Note that $V = V_1 \cup V_2$, where the sets V_1 and V_2 are disjoint, and $E = \big(\bigcup_{i \in V_1} E_1(i)\big) \cup E_2$, where the sets $E_1(i)$, $i \in V_1$, and E_2 are pairwise disjoint. Also note that each of the sets $E_1(i)$, $i \in V_1$, consists of just two edges. To ease the exposition, we will assume that any IP POP in the distributed network is eligible for dismantling. (The model is easily adjusted to account for the case that only some subset of the IP POPs are eligible for dismantling.) Furthermore, we will only allow an IP POP to be dismantled, if any other IP POP using it as a transit node is also dismantled.

If some IP POP is to be dismantled, the customers in the corresponding region must be connected to the network through an alternative IP POP. The following sets specify how customer connections may be transferred between IP POPs in neighboring regions.

- $N(i)$: The set of regions corresponding to IP POPs to which customers in region i can be connected if the IP POP in region i is dismantled ($i \in V$).
- $\overline{N}(i)$: The set of regions from which customers may be connected to the IP POP in region i ($i \in V$).

Note that $N(i), \overline{N}(i) \subseteq V$ and that we have $i \in \overline{N}(i)$ but $i \notin N(i)$ for all $i \in V$.

Example 10.1.1. The network in Figure 10.1 may be divided into a distributed network with node set $V_1 = \{1, 2, 3, 4, 5\}$ and a core network with node set $V_2 = \{6, 7, 8, 9\}$. The corresponding edge sets are given by $E_1(1) = \big\{\{1, 6\}, \{1, 7\}\big\}$, $E_1(2) = \big\{\{2, 6\}, \{2, 7\}\big\}$, $E_1(3) = \big\{\{3, 7\}, \{3, 9\}\big\}$, $E_1(4) = \big\{\{4, 5\}, \{4, 9\}\big\}$, and $E_1(5) = \big\{\{5, 8\}, \{5, 9\}\big\}$ for the distributed network, and $E_2 = \big\{\{6, 7\}, \{6, 8\}, \{6, 9\}, \{7, 9\}, \{8, 9\}\big\}$ for the core network. Clearly, the definition of the sets $N(i)$ and $\overline{N}(i)$, $i \in V$, depends on the practical possibilities to connect customers to IP POPs, and also on any specific preferences of the network operator. As mentioned above, we assume that the network operator considers all IP POPs in the distributed network eligible for dismantling. Suppose now that for practical reasons we have $N(1) = \{2, 6, 7\}$, $N(2) = \{1, 6, 7\}$, $N(3) = \{7, 9\}$, $N(4) = (5, 9)$, and $N(5) = \{4, 8\}$. Also, since the IP POPs in the core network cannot be dismantled, we let $N(6) = N(7) = N(8) = N(9) = \emptyset$. Now, the sets $\overline{N}(i)$, $i \in V$, should be defined consistently by $\overline{N}(1) = \overline{N}(2) = \{1, 2\}$, $\overline{N}(3) = \{3\}$, $\overline{N}(4) = \overline{N}(5) = \{4, 5\}$, $\overline{N}(6) = \{1, 2, 6\}$,

$\overline{N}(7) = \{1, 2, 3, 7\}$, $\overline{N}(8) = \{5, 8\}$, and $\overline{N}(9) = \{3, 4, 9\}$. Finally, note that IP POP 5 is used as a transit node by IP POP 4, and hence it can only be dismantled if IP POP 4 is dismantled.

10.1.3 Variables

The most important group of decisions to be made is whether each individual IP POP should be dismantled, or maintained and possibly upgraded. To this end we assume that a set H of different IP POP classes are available, each class $h \in H$ being characterized by a certain customer- and switch-capacity of the IP POP, and the class $0 \in H$ corresponding to dismantling of the IP POP. For each region $i \in V$ we denote by $H(i) \subseteq H$ the set of available IP POP classes that may be selected in region i. The dimensioning of IP POPs is now described by the variables

$$\cdot\ x_{ih} = \begin{cases} 1 & \text{if a class } h \text{ IP POP is selected in region } i\ (i \in V,\ h \in H(i)); \\ 0 & \text{otherwise.} \end{cases}$$

The next group of decisions concerns the connection of customers to the network. We assume that all customers in regions where the IP POP is maintained remain connected to the network through that particular IP POP. Customers in regions where the IP POP is to be dismantled, however, must be connected to the network through an alternative IP POP. The transfer of customer connections to alternative IP POPs clearly cannot be decided on an individual basis, and hence we divide the customers in any particular region into a number of groups, so that, if the IP POP in that region is to be dismantled, all customers in a given group must be connected to the network through the same alternative IP POP. For each region $i \in V$, we denote by $G(i)$ the set of customer groups in region i. Note that the sets $G(i)$, $i \in V$, are disjoint and hence form a partition of the set of all customer groups, $G = \bigcup_{i \in V} G(i)$. Now, the connection of customers to the network is given by the variables

$$\cdot\ y_{ig} = \begin{cases} 1 & \text{if group } g \text{ is connected to IP POP in region } i\ \left(i \in V,\ g \in \bigcup_{j \in \overline{N}(i)} G(j)\right); \\ 0 & \text{otherwise.} \end{cases}$$

The last group of decisions concerns the dimensioning of transmission links. As previously discussed, we assume that all IP POPs in the distributed network are connected to the rest of the network through two alternatively conveyed transmission links of equal type and capacity, and hence we use just one variable to represent the dimensioning of these two connections for each IP POP. The transmission links in the distributed network currently use either E1 circuits or ATM PVCs, and we allow a future change of type for each IP POP. Also, the transmission links from an IP POP in the distributed network may be replaced by STM-1's (implying that the IP POP becomes part of the future core network). Thus the connections in the distributed network may be selected to be one of three types — E1 circuits (type 1), ATM PVCs (type 2), or STM-1 (type 3). If STM-1 connections are selected, the standard capacity of 155 Mbit/s is provided on each of the two connections. If, on the other hand, E1 circuits or ATM PVCs are selected, the capacity of the transmission links must be decided. Hence the variables concerning dimensioning of transmission links in the distributed network are

$\cdot\ z_{il} = \begin{cases} 1 & \text{if type } l \text{ connections are selected from region } i\ (i \in V_1,\ l = 1,2,3); \\ 0 & \text{otherwise.} \end{cases}$

$\cdot\ v_i :$ Number of E1 circuits to be added to both connections from region i $(i \in V_1)$.

$\cdot\ w_i :$ ATM PVC capacity to be added to both connections from region i $(i \in V_1)$.

For the connections in the core network, three types are available — STM-1 (type 1), STM-4 (type 2) or STM-16 (type 3) — and hence we only have one group of decisions,

$\cdot\ u_{ijl} = \begin{cases} 1 & \text{if the connection } \{i,j\} \text{ is selected of type } l\ (\{i,j\} \in E_2,\ l = 1,2,3); \\ 0 & \text{otherwise.} \end{cases}$

10.1.4 Parameters

Associated with each group of decisions is a corresponding cost term. First we have the cost associated with the selection of a certain IP POP class for each region and the cost of connecting customers to the future network.

$\cdot\ p_{ih} :$ Cost of selecting a class h IP POP in region i $(i \in V,\ h \in H(i))$. This parameter includes all costs associated with any potential upgrading of the IP POP in region $i \in V$, as well as the expected present value of future maintenance costs. The cost of upgrading an IP POP is given as the cost of new equipment minus the value of existing equipment. Thus we note in particular that for all regions $i \in V$ such that $0 \in H(i)$ we have $p_{i0} \leq 0$.

$\cdot\ q_{ig} :$ Cost of connecting group g to the IP POP in region i $\big(i \in V,\ g \in \bigcup_{j \in \overline{N}(i)} G(j)\big)$.

Next is the cost associated with capacity installments on the links of the network. For the transmission links in the distributed network, the cost structure is rather complicated since the three types of capacity — E1 circuits, ATM PVCs, or STM-1 — are completely different in nature. If E1 circuits are preferred from some particular IP POP, capacity is installed in lumps of 2 Mbit/s on each of the two links connecting this IP POP to the rest of the network. If, on the other hand, ATM PVCs are preferred, a fixed cost of the ATM equipment is incurred, whereas the cost of increasing capacity on each of the two connections is assumed to be linear. (Obviously, capacity is also installed in lumps on ATM connections, but the assumption of linear cost in this model, is justified by the fact that ATM connections are shared by the IP network with a number of other services.) Finally, a fixed cost is incurred when connecting an IP POP to the rest of the network by two STM-1 connections, each providing the capacity of 155 Mbit/s. The structure of the capacity expansion cost for a transmission link, on which no capacity is currently installed, is illustrated in Figure 10.2. The matter is further complicated by the fact that some capacity, in the form of either E1 circuits or ATM PVCs, is already installed on all transmission links in the distributed network. If an IP POP in the distributed network is pointed out for dismantling or if a change of type of capacity on the connection from the IP POP to the rest of the network is decided, some of the currently installed equipment may be reused and hence represents a certain value. If ATM PVCs are currently used,

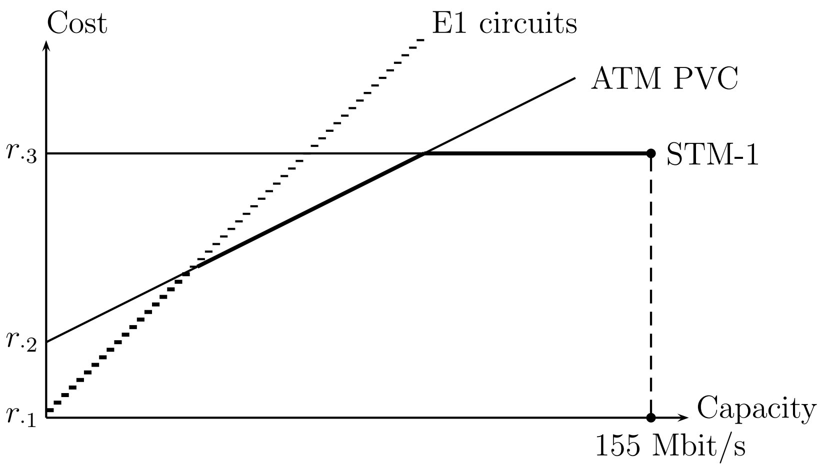

Figure 10.2: Cost structure for transmission links in the distributed network.

the reusable "equipment" consists of ATM equipment installed in the IP POP as well as the ATM PVC capacity currently used on the connections from the IP POP to the rest of the network. If E1 circuits are currently used, the only reusable equipment is the actual circuits. All in all, the following parameters will be used to describe the capacity expansion of transmission links in the distributed network.

- r_{il} : Fixed cost incurred if type l connections are selected from the IP POP in region i ($i \in V_1,\ l = 1, 2, 3$). Note that if type l connections are currently used from the IP POP in region i, we have $r_{il} = 0$. If, on the other hand, some other type of connections are currently used, the parameter r_{il} represents the fixed cost associated with type l connections minus the value of existing equipment that may be reused.

- a_i : Cost of adding an E1 circuit on both connections from the IP POP in region i to the rest of the network ($i \in V_1$).

- b_i : Cost of increasing the ATM PVC capacity by 1 Mbit/s on both connections from the IP POP in region i to the rest of the network ($i \in V_1$).

- $\tilde{v}_i$: The number of E1 circuits currently installed on each of the two connections from the IP POP in region i to the rest of the network ($i \in V_1$).

- $\tilde{w}_i$: The current ATM PVC capacity of each of the two connections from the IP POP in region i to the rest of the network ($i \in V_1$).

For the transmission links in the core network, on the other hand, the cost structure is significantly simpler, since a fixed cost is associated with the capacity provided by the preferred type of connection.

- c_{ijl} : Cost incurred if a type l connection is selected between the IP POPs in regions i and j ($\{i, j\} \in E_2,\ l = 1, 2, 3$). Again, the parameter is a net cost given as the cost of new equipment minus the value of existing equipment.

- C_l : Capacity of a type l connection in the core network ($l = 1, 2, 3$).

The next group of parameters concerns the dimensioning of IP POPs. As previously mentioned, each IP POP class is characterized by a certain customer- and switch-capacity. The customer-capacity restricts the number of customers that can be connected to the network through an IP POP, and is expressed as a number of sockets available for customer connections. The switch-capacity, on the other hand, restricts the amount of traffic that can be switched by an IP POP and is measured in Mbit/s.

· M_h : Customer-capacity of a class h IP POP ($h \in H$). Note that $M_0 = 0$.

· N_h : Switch-capacity of a class h IP POP ($h \in H$). Note that $N_0 = 0$.

The final group of parameters describes demand in the form of requests for customer connections and IP traffic, the latter being modeled as in previous chapters by means of a set K of commodities.

Remark 10.1.2. As pointed out also in previous chapters, the set K of commodities may essentially be defined in two different ways. In the present context, a disaggregated formulation defines each commodity $k \in K$ as traffic from some customer group $o(k) \in G$ to another group $d(k) \in G$, thus resulting in a total of $O(|G|^2)$ commodities. An aggregated formulation, on the other hand, defines each commodity $k \in K$ as all traffic originating in a given group $o(k) \in G$, thus resulting in a total of only $O(|G|)$ commodities. In general, the disaggregated formulation provides a more detailed description of traffic, favorable for example when survivability requirements are to be formulated. The aggregated formulation, on the other hand, provides the advantage of reducing considerably the number of variables and constraints.

As already mentioned, future demand is not known with certainty at the point in time when the network design problem is to be solved, and we include this inherent uncertainty in the formulation using a scenario approach. Hence a number of scenarios is defined, each scenario corresponding to a possible future outcome of random demand.

· L_g^s : Number of sockets required to connect group g to an IP POP under scenario s ($g \in G,\ s \in \{1, \ldots, S\}$).

· D_{kg}^s : Net demand for commodity k from group g under scenario s ($k \in K,\ g \in G,\ s \in \{1, \ldots, S\}$). We emphasize that this parameter represents net demand, and hence it is given a sign so that for all $k \in K$ and $s \in \{1, \ldots, S\}$ we have $\sum_{g \in G} D_{kg}^s = 0$, $D_{kg}^s \geq 0$ for $g \in G \setminus \{o(k)\}$, and $D_{k,o(k)}^s < 0$.

· d_g^s : Total amount of traffic that terminates at group g under scenario s ($g \in G$, $s \in \{1, \ldots, S\}$). The parameter includes the traffic to group g from any other customer group under scenario s (i.e. $\sum_{k \in K: o(k) \neq g} D_{kg}^s$) as well as all internal traffic in group g under scenario s.

Remark 10.1.3. Employing the aggregated formulation of commodities, we see that for $s \in \{1, \ldots, S\}$, $k \in K$, and $g \in G \backslash \{o(k)\}$, the parameter D_{kg}^s represents traffic from group $o(k)$ to group g, whereas the parameter $D_{k,o(k)}^s$ represents total traffic from group $o(k)$ to all other customer groups. Using the disaggregated formulation, on the other hand, we see that for $s \in \{1, \ldots, S\}$ and $k \in K$, the parameters $D_{k,o(k)}^s$ and $D_{k,d(k)}^s$ both represent traffic from group $o(k)$ to group $d(k)$, whereas $D_{kg}^s = 0$ for $g \in G \setminus \{o(k), d(k)\}$.

Finally, we will need to define the maximum aggregate traffic demand

$$\cdot\ D = \max_{1 \le s \le S} \Big\{ \sum_{k \in K} \sum_{g \in G \setminus \{o(k)\}} D^s_{kg} \Big\}.$$

10.1.5 Capacity Constraints

Capacity constraints will be imposed in two different contexts. The first group of capacity constraints concern the customer-capacity of each individual IP POP. Here we find it convenient to introduce for $i \in V$ and $s \in \{1, \ldots, S\}$ the variable λ^s_i representing the shortage of sockets for customer connections to the IP POP in region i under scenario s. For each scenario $s \in \{1, \ldots, S\}$ the constraints are now formulated as

$$\sum_{h \in H(i)} M_h x_{ih} + \lambda^s_i \ge \sum_{j \in \overline{N}(i)} \sum_{g \in G(j)} L^s_g y_{ig}, \quad i \in V. \tag{10.1.1}$$

The remaining capacity constraints concern the restrictions on the flow of IP traffic in the network. To formulate these constraints, we need to determine the traffic flow in the network under any scenario. To this end we define for each $s \in \{1, \ldots, S\}$, $\{i, j\} \in E$ and $k \in K$ the variables f^s_{ijk} and f^s_{jik} representing the flow under scenario s of commodity k on the edge $\{i, j\}$ in direction from i to j and j to i, respectively. The flow of traffic is now determined by the following flow conservation constraints, stating that the net flow of a commodity into some IP POP should equal the net demand for the commodity from customers connected to that particular IP POP. Hence for each scenario $s \in \{1, \ldots, S\}$ we impose the constraints

$$\sum_{j:\{i,j\} \in E} f^s_{jik} - \sum_{j:\{i,j\} \in E} f^s_{ijk} = \sum_{j \in \overline{N}(i)} \sum_{g \in G(j)} D^s_{kg} y_{ig}, \quad i \in V,\ k \in K. \tag{10.1.2}$$

Remark 10.1.4. Note that the flow conservation constraints (10.1.2) correspond to those of a standard multicommodity flow problem, and in particular that the possibility of free flow distribution is implicitly assumed. In other words, we implicitly assume that the flow of traffic between any pair of nodes may be divided arbitrarily among a number of different paths. This is not, however, in accordance with the facts of IP routing cf. the discussion in Holmberg and Yuan [62]. Still, computational results presented by Holmberg and Yuan show that the standard multicommodity flow constraints provide a reasonable approximation of a much more complex model for IP routing.

The following constraints concern the switch-capacity of the IP POPs. Again we define for $i \in V$ and $s \in \{1, \ldots, S\}$ the variable γ^s_i representing the shortage of switch-capacity of the IP POP in region i under scenario s. Hence for $s \in \{1, \ldots, S\}$ the constraints are formulated as

$$\sum_{h \in H(i)} N_h x_{ih} + \gamma^s_i \ge \sum_{j \in \overline{N}(i)} \sum_{g \in G(j)} d^s_g y_{ig} + \sum_{k \in K} \sum_{j:\{i,j\} \in E} f^s_{ijk}, \quad i \in V. \tag{10.1.3}$$

Here the amount of traffic switched by an IP POP is determined as the total flow out of the IP POP — that is, the sum of traffic terminating at customers connected to the IP POP and traffic sent on through the network.

For the transmission links in the distributed network, we require not only that the total traffic in either direction on a link should not exceed the capacity of that link, but also that each of the two alternative connections from an IP POP in the distributed network to the rest of the network, has enough capacity to carry 60% of the total traffic into and out of the IP POP. Here we define for $i \in V_1$ and $s \in \{1, \ldots, S\}$ the variables τ_i^s representing the shortage of capacity under scenario s on the two transmission links from the IP POP in region i to the rest of the network. Now, for $s \in \{1, \ldots, S\}$ the constraints are,

$$2(\tilde{v}_i z_{i1} + v_i) + \tilde{w}_i z_{i2} + w_i + C_1 z_{i3} + \tau_i^s \geq \sum_{k \in K} f_{ijk}^s, \quad i \in V_1,\ \{i,j\} \in E_1(i), \tag{10.1.4a}$$

$$2(\tilde{v}_i z_{i1} + v_i) + \tilde{w}_i z_{i2} + w_i + C_1 z_{i3} + \tau_i^s \geq \sum_{k \in K} f_{jik}^s, \quad i \in V_1,\ \{i,j\} \in E_1(i), \tag{10.1.4b}$$

$$2(\tilde{v}_i z_{i1} + v_i) + \tilde{w}_i z_{i2} + w_i + C_1 z_{i3} + \tau_i^s \geq 0.6 \sum_{k \in K} \sum_{j:\{i,j\} \in E_1(i)} f_{ijk}^s, \quad i \in V_1, \tag{10.1.4c}$$

$$2(\tilde{v}_i z_{i1} + v_i) + \tilde{w}_i z_{i2} + w_i + C_1 z_{i3} + \tau_i^s \geq 0.6 \sum_{k \in K} \sum_{j:\{i,j\} \in E_1(i)} f_{jik}^s, \quad i \in V_1. \tag{10.1.4d}$$

Remark 10.1.5. Note that since we are considering an IP network using optical transmission systems, each link in the network can carry flow in either direction and, more importantly, these flows do not interfere.

Finally, for the transmission links in the core network, the only capacity constraints are for each scenario $s \in \{1, \ldots, S\}$ that

$$\sum_{l=1}^{3} C_l u_{ijl} + \sigma_{ij}^s \geq \sum_{k \in K} f_{ijk}^s, \quad \{i,j\} \in E_2, \tag{10.1.5a}$$

$$\sum_{l=1}^{3} C_l u_{ijl} + \sigma_{ij}^s \geq \sum_{k \in K} f_{jik}^s, \quad \{i,j\} \in E_2, \tag{10.1.5b}$$

where σ_{ij}^s denotes the shortage of capacity on the transmission link $\{i,j\} \in E_2$ under scenario s.

10.1.6 Two-Stage Formulation

As previously discussed, the decisions concerning network design must be made so as to minimize total costs incurred while ensuring that enough capacity is installed to accommodate any future demand scenario. The latter restriction is formulated implicitly using functions Q^s, defined for each scenario $s \in \{1, \ldots, S\}$ by

$$\begin{aligned} Q^s(x,y,z,v,w,u) = \min & \sum_{i \in V} \left(\lambda_i^s + \gamma_i^s\right) + \sum_{i \in V_1} \tau_i^s + \sum_{\{i,j\} \in E} \sigma_{ij}^s \\ \text{s.t.} \ & (10.1.1) - (10.1.5), \\ & f^s \in \mathbb{R}_+^{2|E||K|},\ \lambda^s, \gamma^s \in \mathbb{R}_+^{|V|},\ \tau^s \in \mathbb{R}_+^{|V_1|},\ \sigma^s \in \mathbb{R}_+^{|E_2|}. \end{aligned} \tag{10.1.6}$$

Obviously, sufficient capacity to accommodate demand scenario $s \in \{1, \dots, S\}$ is installed if and only if the decisions concerning network design are such that $Q^s(x, y, z, v, w, u) = 0$. Hence the network design problem may be formulated as

$$\min \sum_{i \in V} \sum_{h \in H(i)} p_{ih} x_{ih} + \sum_{i \in V} \sum_{j \in \overline{N}(i)} \sum_{g \in G(j)} q_{ig} y_{ig} + \sum_{i \in V_1} \Big(\sum_{l=1}^{3} r_{il} z_{il} + a_i v_i + b_i w_i \Big) + \sum_{\{i,j\} \in E} \sum_{l=1}^{3} c_{ijl} u_{ijl} \tag{10.1.7a}$$

$$\text{s.t.} \sum_{h \in H(i)} x_{ih} = 1 \qquad i \in V \tag{10.1.7b}$$

$$x_{j0} \le x_{i0} \qquad i, j \in V_1,\ \{i, j\} \in E_1(i) \tag{10.1.7c}$$

$$\sum_{j \in N(i)} y_{jg} = x_{i0} \qquad i \in V_1,\ g \in G(i) \tag{10.1.7d}$$

$$y_{ig} = 1 - x_{i0} \qquad i \in V_1,\ g \in G(i) \tag{10.1.7e}$$

$$y_{ig} = 1 \qquad i \in V_2,\ g \in G(i) \tag{10.1.7f}$$

$$\sum_{l=1}^{3} z_{il} = 1 - x_{i0} \qquad i \in V_1 \tag{10.1.7g}$$

$$v_i \le D z_{i1} \qquad i \in V_1 \tag{10.1.7h}$$

$$w_i \le D z_{i2} \qquad i \in V_1 \tag{10.1.7i}$$

$$\sum_{l=1}^{3} u_{ijl} = 1 \qquad \{i, j\} \in E_2 \tag{10.1.7j}$$

$$Q^s(x, y, z, v, w, u) = 0 \qquad s = 1, \dots, S \tag{10.1.7k}$$

$$x \in \mathbb{B}^X,\ y \in \mathbb{B}^Y,\ z \in \mathbb{B}^{3|V_1|}, u \in \mathbb{B}^{3|E_2|},\ v \in \mathbb{Z}_+^{|V_1|},\ w \in \mathbb{R}_+^{|V_1|}. \tag{10.1.7l}$$

Here $X = \sum_{i \in V} |H(i)|$ and $Y = \sum_{i \in V} \sum_{j \in \overline{N}(i)} |G(j)|$. The objective function (10.1.7a) consists of four terms, corresponding to installment of IP POPs, connection of customers to the network, capacity expansion of transmission links in the distributed network, and capacity expansion of transmission links in the core network, respectively. According to (10.1.7b), one IP POP class is selected for each region, and (10.1.7c) ensures that an IP POP is only dismantled if any other IP POP using it as a transit node is also dismantled. If some IP POP in the distributed network is to be dismantled ($x_{\cdot 0} = 1$), the customers in the corresponding region should be connected to the network through an alternative IP POP. If, on the other hand, the IP POP is maintained ($x_{\cdot 0} = 0$), the customers in the corresponding region remain connected to the network through this particular IP POP. This is achieved by the constraints (10.1.7d) and (10.1.7e). For an IP POP in the core network, on the other hand, all customers in the corresponding region remain connected to the network through this particular IP POP cf. (10.1.7f). Next, a type of connection should be selected from each IP POP in the distributed network that is maintained cf. (10.1.7g). Also, for each transmission link connecting an IP POP in the distributed network to the rest of the network, (10.1.7h) and

(10.1.7i) ensure that capacity in the form of E1 circuits or ATM PVCs may be expanded if and only if this particular type of capacity is selected. (Note that we use the maximum aggregate traffic demand D as a "big-M coefficient".) Next, the constraint (10.1.7j) requires a type of connection to be selected for each transmission link connecting a pair of IP POPs in the core network. Finally, (10.1.7k) ensures that enough capacity is installed to accommodate all demand scenarios.

Remark 10.1.6. Clearly, no customers should be connected to the network through an IP POP that is pointed out for dismantling. Hence any feasible solution of the network design problem should satisfy the constraints

$$y_{ig} \leq 1 - x_{i0} \quad i \in V_1,\ g \in \textstyle\bigcup_{j\in\overline{N}(i)} G(j).$$

These constraints are implied, however, by e.g. the constraints (10.1.1) and (10.1.7k), and hence we do not include them in the formulation.

Remark 10.1.7. The second-stage constraint (10.1.3) ensures that no flow of traffic can occur out of an IP POP that is pointed out for dismantling. Hence if the IP POP in some region is to be dismantled, the constraint (10.1.2) ensures that no flow of traffic can occur into this IP POP since no customer groups are connected to it cf. Remark 10.1.6. Thus no flow of traffic can occur into or out of IP POPs that are to be dismantled.

10.2 Solution Procedure

To ease the exposition we find it convenient in the following to simplify the notation, writing the aggregate first-stage solution vector (x, y, z, u, v, w) simply as $\tilde{x}$, and for each scenario $s \in \{1, \ldots, S\}$ writing the vector of second-stage shortage variables $(\lambda^s, \gamma^s, \tau^s, \sigma^s)$ simply as ρ^s. Also, we will consider the network design problem (10.1.7) only in the conceptual form,

$$\begin{aligned} \min\ & c\tilde{x} \\ \text{s.t. } & A\tilde{x} = b, \\ & Q^s(\tilde{x}) = 0, \quad s = 1, \ldots, S, \\ & \tilde{x} \in \tilde{X}, \end{aligned} \tag{10.2.1}$$

where $c \in \mathbb{R}^{n_1}$ represents the first-stage objective (10.1.7a), $b \in \mathbb{R}^{m_1}$ and $A \in \mathbb{R}^{m_1 \times n_1}$ represent the first-stage constraints (10.1.7b)-(10.1.7j), and $\tilde{X} \subseteq \mathbb{R}^{n_1}_+$ is a subset, restricting the appropriate components of $\tilde{x}$ to be either binary, integer or real numbers. Likewise, the second-stage problem (10.1.6) is considered only in a conceptual form, written for $\tilde{x} \in \mathbb{R}^{n_1}$ and $s \in \{1, \ldots, S\}$ as

$$Q^s(\tilde{x}) = \min\{e\rho^s \mid W^s f^s + W'\rho^s \geq h^s - T^s\tilde{x},\ f^s \in \mathbb{R}^{n_2}_+,\ \rho^s \in \mathbb{R}^{n_2'}_+\}, \tag{10.2.2}$$

where $h^s \in \mathbb{R}^{m_2}$, $W^s \in \mathbb{R}^{m_2 \times n_2}$, $W' \in \mathbb{R}^{m_2 \times n_2'}$, and $T^s \in \mathbb{R}^{m_2 \times n_1}$ represent the capacity constraints (10.1.1)-(10.1.5), and $e = (1, \ldots, 1) \in \mathbb{R}^{n_2'}$.

The fundamental idea in the cutting plane method for problem (10.2.1) presented below, is to relax the constraints $Q^s(\tilde{x}) = 0$, $s = 1, \ldots, S$, and iteratively re-enforce them by means of so-called feasibility cuts. Hence, we start with a relaxation of problem (10.2.1), referred to as the master problem, in which the constraints $Q^s(\tilde{x}) = 0$, $s = 1, \dots, S$, have been removed. Given a solution $\tilde{x}^\nu \in \mathbb{R}^{n_1}$ of the master problem in some iteration ν, the feasibility cuts are derived from the second-stage problem (10.2.2) with $\tilde{x} = \tilde{x}^\nu$. Specifically, we consider the corresponding dual problem, defined for $s \in \{1, \dots, S\}$ by

$$Q^s(\tilde{x}) = \max\big\{(h^s - T^s\tilde{x})\pi^s \mid \pi^s W^s \le 0,\ \pi^s W' \le e,\ \pi^s \in \mathbb{R}^{m_2}_+\big\}. \qquad (10.2.3)$$

Obviously, problems (10.2.2) and (10.2.3) are both feasible, and hence they are both solvable and their optimal values are identical and clearly non-negative. In particular, if $\tilde{x}^\nu \in \mathbb{R}^{n_1}$ is such that insufficient capacity is installed to accommodate some demand scenario $s \in \{1, \dots, S\}$ (i.e. $Q^s(\tilde{x}^\nu) > 0$), then a feasible solution $\pi^{s,\nu}$ of the dual problem (10.2.3) exists such that $(h^s - T^s\tilde{x}^\nu)\pi^{s,\nu} > 0$. Moreover, since $\pi^{s,\nu}$ is feasible for the dual problem (10.2.3), we see that for all $\tilde{x} \in \mathbb{R}^{n_1}$ with $Q^s(\tilde{x}) = 0$ we have

$$(h^s - T^s\tilde{x})\pi^{s,\nu} \le 0. \qquad (10.2.4)$$

The constraint (10.2.4) is referred to as a feasibility cut, and as described above it can be used to cut off the current solution of the master problem $\tilde{x}^\nu \in \mathbb{R}^{n_1}$ whenever $Q^s(\tilde{x}^\nu) > 0$ for some scenario $s \in \{1, \dots, S\}$.

The algorithm progresses by sequentially solving a master problem and adding violated feasibility cuts generated through the solution of subproblems (10.2.2)-(10.2.3).

Algorithm 10.1

Step 1 (*Initialization*) Set $\nu = 0$, and let the current master problem be defined by $\min\big\{c\tilde{x} \mid A\tilde{x} = b,\ \tilde{x} \in \tilde{X}\big\}$.

Step 2 (*Solve master problem*) Set $\nu = \nu + 1$. Solve the current master problem and let $\tilde{x}^\nu$ be an optimal solution vector.

Step 3 (*Solve subproblems*) For each $s \in \{1, \dots, S\}$, solve the second-stage problem (10.2.2) with $\tilde{x} = \tilde{x}^\nu$, and let $\pi^{s,\nu}$ be a corresponding optimal dual solution. If $(h^s - T^s\tilde{x}^\nu)\pi^{s,\nu} > 0$ for some $s \in \{1, \dots, S\}$, add a feasibility cut (10.2.4) to the master problem and return to Step 2. Otherwise, stop; the current solution $\tilde{x}^\nu$ is optimal.

Proposition 10.2.1. *If problem (10.2.1) is feasible, then Algorithm 10.1 terminates with an optimal solution of the problem in a finite number of iterations.*

Proof. Finite convergence is an immediate consequence of validity of the feasibility cuts cf. the discussion above, and the fact that only a finite number of extreme points of the feasible region in (10.2.3) exist. □

Remark 10.2.1. The requirement of feasibility in Proposition 10.2.1 is certainly not unreasonable, since it should always be possible to install sufficient capacity to accommodate demand. Clearly, though, problem (10.1.7) becomes infeasible if demand rises beyond a certain level. Therefore, in such a case, the formulation is not appropriate, and one may have to allow for the placement of new IP POPs as well as for the installment of several facilities (STM-1, STM-4 or STM-16) on connections in the future core network. For the IP network of TDC, however, the formulation presented here was found appropriate.

Remark 10.2.2. Note that Algorithm 10.1 employs a mixed-integer programming formulation of the master problem in each iteration. The generation of feasibility cuts, however, may as well be carried out for fractional solutions cf. our discussion above, and hence it seems natural to assume that it is not worthwhile to put a lot of effort into finding integral first-stage solutions in early iterations of the algorithm. In fact, we recall that the capacitated network design problem was formulated as a two-stage stochastic program with linear recourse in Chapter 7, and a solution method was proposed that is similar in vein to Algorithm 10.1, but in which integer requirements are initially removed in the master problem. This algorithm then proceeds to restore integrality and feasibility simultaneously through a branch-and-cut scheme, with feasibility cuts being generated at all nodes of the branching tree. Furthermore, as pointed out also in Remark 7.3.3 on page 103, Albareda-Sambola, van der Vlerk, and Ferńandez [2], compared different versions of a similar algorithm for a class of stochastic generalized assignment problems, and concluded that such a branch-and-cut scheme performed superior to a branch-first-cut-second scheme such as Algorithm 10.1. We did in fact also try a branch-and-cut algorithm for problem (10.2.1). It turned out, however, that for the particular instance considered here, the major effort lies in solving the second-stage problems, whereas solving even a mixed-integer formulation of the master problem is relatively easily done with the CPLEX Mixed Integer optimizer. This means that generating cuts via the solution of the second-stage problems throughout the branching process is simply too time consuming, and very little movement in the lower bound was observed for the branch-and-cut algorithm. Hence for this problem, the branch-first-cut-second scheme of Algorithm 10.1 actually proved superior.

10.2.1 Valid Inequalities

As pointed out in Remark 10.2.2, for the IP network of TDC the main computational effort in solving problem (10.1.7) using Algorithm 10.1, lay in the generation of feasibility cuts. Furthermore, since no capacity constraints are initially present in the master problem, a direct application of Algorithm 10.1 as it was presented above would require a large number of feasibility cuts to be imposed to properly reflect the capacity requirements in the second stage. In fact we tried a direct application of Algorithm 10.1, and observed that a vast amount of time was spent solving second-stage problems to generate feasibility cuts, achieving only very little movement in the optimal value of the master problem. Therefore, to improve performance of the algorithm, we determined a large number of capacity constraints that could be generated a priori without solving any second-stage problems. In the following we choose to work with an aggregated formulation of the commodities, defining a commodity for each customer group, i.e. $K = G$, so that commodity $k \in K$ corresponds to IP traffic originating at group k. If a disaggregated formulation is employed, however, the constraints remain valid with only notational corrections.

First, it is obvious that the constraints concerning the customer-capacity of each individual IP POP can be used directly in the master problem, since they are independent of the routing of traffic in the second stage cf. (10.1.1). Hence we used the constraints

$$\sum_{h \in H(i)} M_h x_{ih} \geq \sum_{j \in \overline{N}(i)} \sum_{g \in G(j)} L_g^s y_{ig}, \quad i \in V,\ s = 1, \ldots, S. \tag{10.2.5}$$

The constraints concerning the switch-capacity of each individual IP POP, on the other hand, clearly depend on the routing of traffic in the second stage cf. (10.1.3), and hence they cannot be used directly in the master problem. Instead, we used the following group of constraints to reflect the second-stage requirement for switch-capacity,

$$\sum_{h\in H(i)} N_h x_{ih} \geq \sum_{j\in\overline{N}(i)} \sum_{g\in G(j)} \left(d_g^s - \sum_{j'\in\overline{N}(i)} \sum_{g'\in G(j')} D_{gg'}^s \right) y_{ig}, \quad i \in V, \ s = 1,\ldots,S. \quad (10.2.6)$$

To see that these are in fact valid inequalities, we note that for any $s \in \{1,\ldots,S\}$, $i \in V$, $j \in \overline{N}(i)$, and $g \in G(j)$ the first term in the parentheses on the right-hand side, $d_g^s \geq 0$, gives the total amount of traffic that terminates at group g under scenario s, whereas the second term, $-\sum_{j'\in\overline{N}(i)} \sum_{g'\in G(j')} D_{gg'}^s \geq 0$, gives the total amount of traffic originating at group g under scenario s (i.e. $-D_{gg}^s$) minus the part of this traffic that terminates at groups that may be connected to the network through the IP POP in region i $\left(\text{i.e. } \sum_{j'\in\overline{N}(i)} \sum_{g'\in G(j')\setminus\{g\}} D_{gg'}^s\right)$. Hence, for any possible allocation of customer groups to IP POPs, the right-hand side of (10.2.6) provides a lower bound on the total amount of traffic that must be switched by the IP POP in region $i \in V$ under scenario $s \in \{1,\ldots,S\}$, and hence it is a valid inequality. In general, the lower bound on the required switch-capacity provided by (10.2.6) is not tight, and in particular we note that no transit traffic is included. Since the switch-capacities of the different IP POP classes are generally far apart, though, the constraints turned out to be quite effective in our computational experiments.

To generate cuts for the required capacity on links in the distributed network, our starting point was the constraints (10.1.4c) and (10.1.4d), stating that each of the two alternative connections from an IP POP in the distributed network to the rest of the network, is required to have enough capacity to carry 60% of the total traffic into and out of the IP POP. Here we used the constraints,

$$2(\tilde{v}_i z_{i1} + v_i) + \tilde{w}_i z_{i2} + w_i + C_1 z_{i3} \geq 0.6 \sum_{j\in\overline{N}(i)} \sum_{g\in G(j)} \left(\sum_{j'\in V_2} \sum_{g'\in G(j')} D_{gg'}^s \right) y_{ig}, \quad i \in V_1, \ s = 1,\ldots,S. \quad (10.2.7a)$$

$$2(\tilde{v}_i z_{i1} + v_i) + \tilde{w}_i z_{i2} + w_i + C_1 z_{i3} \geq 0.6 \sum_{j\in\overline{N}(i)} \sum_{g\in G(j)} \left(\sum_{j'\in V_2} \sum_{g'\in G(j')} D_{g'g}^s \right) y_{ig}, \quad i \in V_1, \ s = 1,\ldots,S. \quad (10.2.7b)$$

To see that (10.2.7a) is a valid inequality, we note that for any $s \in \{1,\ldots,S\}$, $i \in V_1$, $j \in \overline{N}(i)$, and $g \in G(j)$, the term in the parentheses on the right-hand side gives the total amount of traffic that must be routed from group g to customer groups in regions in the core network under scenario s. Hence for any $i \in V_1$ and $s \in \{1,\ldots,S\}$ and for any possible allocation of customer groups to IP POPs, the right-hand side in (10.2.7a) is clearly a lower bound on the amount of traffic from region i to regions in the core network under scenario s, and hence the inequality is valid. Obviously, a similar observation goes for (10.2.7b) only with the direction of traffic reversed. Again we note that the lower bounds provided by (10.2.7a) and (10.2.7b) are obviously not tight, but since traffic between a region $i \in V_1$ and other regions in the distributed network is typically negligible compared to the traffic between region i and the core network, the inequalities turned out quite useful.

Finally, to generate cuts for the required capacity on links in the core network, we used a generalization of the well-known cutset inequalities, employed for example for the capacitated network design problem considered in Chapter 7. (See Section 7.1.2 for a thorough discussion of these inequalities.) To this end, for some $U \subseteq V_2$ we consider the partition $\pi = (U, V_2 \setminus U)$ of V_2, and let $E_\pi = \big\{\{i,j\} \in E_2 \;\big|\; |\{i,j\} \cap U| = 1\big\}$ be the corresponding cutset. Moreover, since only one customer group was defined for all regions corresponding to IP POPs in the core network, cf. Section 10.3.1 below, we let $G(i) = \{g_i\}$, $i \in V_2$, for ease of notation. Now, we used the following constraints,

$$\sum_{\{i,j\}\in E_\pi} \sum_{l=1}^{3} C_l u_{ijl} \geq \max_{s\in\{1,\dots,S\}} \max\left\{ \sum_{i\in U} \sum_{j\in V_2\setminus U} D^s_{g_i g_j}, \sum_{i\in U} \sum_{j\in V_2\setminus U} D^s_{g_j g_i} \right\}, \qquad \pi = (U, V_2 \setminus U),\ U \subseteq V_2. \tag{10.2.8}$$

It is easily seen that these are in fact valid inequalities since the right-hand side of (10.2.8) is a lower bound on the amount of traffic that must be routed across the cutset E_π. Furthermore, cf. our discussion of the inequalities (10.2.7) above, we note that the traffic between regions in the core network and regions in the distributed network is typically negligible compared to the interregional traffic in the core network, and hence the inequalities proved quite useful.

Remark 10.2.3. Note that since the commodity demands for the IP network of TDC was generated by a gravitational model, cf. Section 10.3.1 below, the traffic matrix was in fact symmetric, and hence in practice we did not have to consider traffic in both directions for the link-capacity constraints as stated here in (10.2.7) and (10.2.8).

Obviously, a potentially large number of constraints may be generated a priori from (10.2.5)-(10.2.8). This is true in particular for the generalized cutset inequalities (10.2.8), and hence we chose to consider only those cutsets corresponding to subsets $U \subseteq V_2$ consisting of one or two IP POPs. Moreover, to control the size of the master problem, we chose to generate all cuts from (10.2.5)-(10.2.8) at initialization of the algorithm and store them in a cutpool. Then, in each iteration of the algorithm, before the second-stage problems are solved to possibly generate violated feasibility cuts, this cutpool is scanned to search for violated capacity constraints. If any violated constraints are found they are included in the master problem (at most 10 at a time), and the problem is re-solved.

10.3 Computational Experiments

The algorithm described in the previous section was implemented in C++ using procedures from the callable library of CPLEX 6.6. In particular, the mixed-integer master problem was solved with the CPLEX Mixed Integer optimizer cf. Remark 10.2.2. A series of computational experiments were performed on the IP network of TDC. In this section we discuss the practical application of the model presented in Section 10.1, and we present results of our computational experiments.

10.3.1 Problem Instance

Let us first consider the IP network of TDC. Here the core network consists of 39 IP POPs interconnected by a total of 70 transmission links. The distributed network, on the other hand, consists of 155 IP POPs, most of which are connected to the rest of the network by two alternatively conveyed links of equal type and capacity as assumed in the model. Some exceptions from the idealized network structure of the model presented in Section 10.1 had to be dealt with, however. First, for some IP POPs in the distributed network, the two alternatively conveyed links, connecting the IP POP to the rest of the network, does not presently have equal capacities. In these cases, we simply used the average of the two as the existing capacity for the model input. Second, for specific reasons, a few IP POPs in the distributed network actually have an extra STM-1 link to the rest of the network. These extra links were included in the model, but no upgrading of the connections were allowed. Finally, a few IP POPs in the distributed network are connected to the rest of the network through "hoops" of two IP POPs. This is best illustrated by a small example.

Example 10.3.1. Figure 10.3 illustrates a "hoop", connecting IP POPs 1 and 2 to the rest of the network through IP POPs A and B.

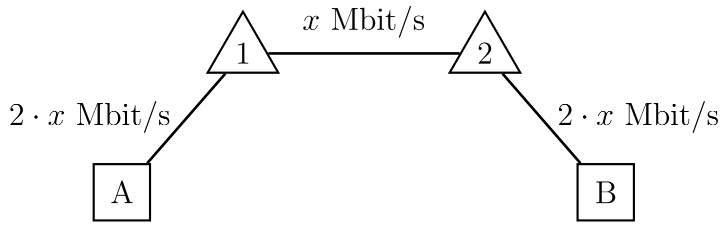

Figure 10.3: Illustration of a "hoop".

Clearly, it is possible to accurately represent such a hoop within the integer programming formulation of the model. We did not find the improved accuracy of such a formulation sufficient to justify the increased model complexity, though, and hence we chose simply to treat IP POPs such as 1 and 2 in Figure 10.3 as if they both had a link with capacity x to IP POP A and a link with capacity x to IP POP B.

A total of seven different IP POP classes were defined (including the class '0'), with at most five potential IP POP classes available for selection in any particular region. Also, for all regions corresponding to IP POPs that may be dismantled, the customers were divided in up to four groups, and up to three potential alternative IP POPs for the customers were specified. (For IP POPs that are not eligible for dismantling, it obviously does not make sense to divide customers into more than one group, or to specify alternative IP POPs.)

All in all, we ended up with a two-stage stochastic program with recourse, containing in the first stage a total of 1960 variables, most of which are binary, and 1137 constraints at initialization. Moreover, when the algorithm progresses, the number of constraints increases as cuts are imposed to re-enforce (10.1.7k). Clearly, though, the special structure of constraints such as e.g. (10.1.7e) and (10.1.7f) allowed CPLEX MIP Presolve to reduce the size of the problem considerably, removing a priori a large number of first-stage variables and constraints. Given a first-stage solution and a particular scenario, the second-stage problem, on the other hand, is a linear programming problem with 206737 continuous variables and 53268 constraints.

The only available demand input for the model was the current groupwise demand for customer connections, denoted here by L_g, $g \in G$, and the current regionwise demand for IP traffic, expressed as the total amount of IP traffic terminating at each IP POP and denoted here by T_i, $i \in V$. Using this data we had to estimate the groupwise demand for IP traffic, and generate a number of future demand scenarios. This was done as follows. First of all, for $s \in \{1, \ldots, S\}$ the future groupwise demand for customer connections under scenario s was calculated as

$$L_g^s = \mu^s \cdot \rho_g^s \cdot L_g, \quad g \in G,$$

where μ^s is a parameter reflecting the average growth in demand for customer connections, and ρ_g^s, $g \in G$, are parameters reflecting regional fluctuations from this average growth. Now, to calculate an estimate of the groupwise demand for IP traffic, we used the estimated future demand for customer connections to split the regionwise demand T_i, $i \in V$, among groups. Hence, for $s \in \{1, \ldots, S\}$ the future groupwise demand for IP traffic under scenario s, expressed as the total volume of IP traffic terminating at each group, was calculated as

$$d_g^s = \lambda^s \cdot \gamma_g^s \cdot \frac{L_g^s}{\sum_{g' \in G(i)} L_{g'}^s} \cdot T_i, \quad i \in V,\ g \in G(i),$$

where λ^s is a parameter reflecting the average growth in demand for IP traffic, and γ_g^s, $g \in G$, are parameters reflecting regional fluctuations from this average growth. The growth factors were all generated by random sampling from appropriate uniform distributions.

Remark 10.3.1. To capture the correlation between growth in demand for customer connections and growth in demand for IP traffic, we actually independently generated parameters μ^s, $\tilde{\lambda}^s$, ρ_g^s and $\tilde{\gamma}_g^s$ for all $g \in G$ and $s \in \{1, \ldots, S\}$, and then defined $\lambda^s = \mu^s \cdot \tilde{\lambda}^s$ and $\gamma_g^s = \rho_g^s \cdot \tilde{\gamma}_g^s$ for $g \in G$ and $s \in \{1, \ldots, S\}$.

Finally, we used an aggregated formulation of the commodities, defining a commodity for each customer group, i.e. $K = G$, so that commodity $k \in K$ corresponds to IP traffic originating at group k. The commodity demand was then calculated by gravitation, using the estimates of the future volume of IP traffic terminating at each group. Hence for $s \in \{1, \ldots, S\}$ the commodity demand for IP traffic was calculated as

$$D_{kg}^s = \frac{d_g^s \cdot d_k^s}{\sum_{g' \in G} d_{g'}^s}, \quad k \in K,\ g \in G \setminus \{k\},$$

and

$$D_{kk}^s = -\sum_{g \in G \setminus \{k\}} D_{kg}^s, \quad k \in K.$$

10.3.2 Computational Results

A series of computational experiments were performed on the IP network of TDC. We generated instances of the problem with 1, 5, 10, 50, and 100 scenarios, and solved the problems using Algorithm 10.1 as described in Section 10.2. The instance with only one scenario was generated by replacing all random parameters by their expected values, and hence it will be referred to as the expected value problem (EVP). At termination of each run we recorded the number of iterations performed, the number of feasibility cuts applied, the total number of generated cuts, the number of cuts in the master problem (referred to as *active cuts*), and the CPU time spent by the procedure. Results are reported in Table 10.1.

Table 10.1: Computational Results

S	Iterations	Feas. cuts	Total cuts	Active cuts	CPU time
(EVP) 1	15	14	1337	152	2:17
5	13	32	3183	680	4:12
10	19	119	5555	1426	7:23
50	20	520	24234	2013	25:42
100	23	1096	47651	2662	47:33

The optimal solution of the two instances with 50 and 100 scenarios, respectively, suggested dismantling of the same particular 18 IP POPs. The optimal solution of the instances with 5 and 10 scenarios only disagreed with this suggestion for one and four IP POPs, respectively. The optimal solution of the expected value problem, on the other hand, suggested dismantling of just 10 IP POPs, one of which was not suggested for dismantling in any of the other solutions. To investigate the effect of using a stochastic programming model with multiple scenarios, we fixed the dismantling of IP POPs suggested by the solution of the expected value problem, and subsequently solved the stochastic programming problem with the same 100 scenarios as before. The resulting total cost turned out to be 3.5% larger than the minimum cost determined in the previous run. Hence, given the size of the total installment cost, the saving obtained by solving the stochastic programming problem rather than the expected value problem, is considerable.

Appendix A

Prerequisites from Probability Theory

In this appendix we present a few basic concepts and results from probability theory that are used throughout the thesis. For a more thorough discussion of these topics we refer to the textbooks by e.g. Dudley [43] and Hoffmann-Jørgensen [61].

A.1 Probability Spaces

Given some set Ω let us recall that the so-called power set 2^Ω is the set of all subsets of Ω. Also, given some set $F \in 2^\Omega$ we denote by $F^c = \Omega \setminus F$ the complement of F. Now, our starting point here is the following definitions.

Definition A.1.1. A non-empty collection of subsets $\mathcal{F} \subseteq 2^\Omega$ of a set Ω is said to be a *σ-algebra* of subsets of Ω if the following two properties hold:

(i) $F^c \in \mathcal{F} \quad \text{whenever} \quad F \in \mathcal{F};$

(ii) $\displaystyle\bigcup_{n=1}^{\infty} F_n \in \mathcal{F} \quad \text{whenever} \quad F_1, F_2, \ldots \in \mathcal{F}.$

Definition A.1.2. A tuple $(\Omega, \mathcal{F})$ is said to be a *measurable space* if Ω is a set, and $\mathcal{F}$ is a σ-algebra of subsets of Ω.

Remark A.1.1. Given a set Ω and a measurable space $(\Lambda, \mathcal{G})$, any mapping $f : \Omega \mapsto \Lambda$ is easily seen to introduce a σ-algebra of subsets of Ω, namely $\mathcal{F} = f^{-1}(\mathcal{G})$. This is referred to as the σ-algebra generated by f.

Definition A.1.3. Let $(\Omega, \mathcal{F})$ be a measurable space. A real-valued function $P : \mathcal{F} \mapsto \mathbb{R}$ is said to be a *probability measure* on $\mathcal{F}$ if the following three properties hold:

(i) $P(\Omega) = 1;$

(ii) $P(F) \geq 0 \quad \text{whenever} \quad F \in \mathcal{F};$

(iii) $\displaystyle P\Big(\bigcup_{n=1}^{\infty} F_n\Big) = \sum_{n=1}^{\infty} P(F_n) \quad \text{whenever} \quad F_1, F_2, \ldots \in \mathcal{F},\ F_n \cap F_m = \emptyset,\ m \neq n.$

With these basic definitions we may now formally define the notion of a probability space.

Definition A.1.4. A tuple $(\Omega, \mathcal{F}, P)$ is said to be a *probability space* if Ω is a set, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and P is a probability measure on $\mathcal{F}$.

Throughout the thesis we are particularly concerned with one specific σ-algebra, the *Borel σ-algebra* on some subset $\Xi \subseteq \mathbb{R}^N$ denoted by $\mathcal{B}(\Xi)$. This σ-algebra may be equivalently defined either as the smallest σ-algebra on Ξ containing all open subsets of Ξ, or as the smallest σ-algebra on Ξ containing all closed subsets of Ξ. A set $B \in \mathcal{B}(\Xi)$ is called a *Borel set* and is said to be *Borel measurable*. Hence, for any subset $\Xi \subseteq \mathbb{R}^N$, $(\Xi, \mathcal{B}(\Xi))$ is a measurable space, and a probability measure on $\mathcal{B}(\Xi)$ will be referred to as a *Borel probability measure* on Ξ. The set of all Borel probability measures on Ξ is denoted by $\mathcal{P}(\Xi)$.

A.2 Random Variables and Random Vectors

To formally define the notion of random variables and random vectors, we need the following.

Definition A.2.1. Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be measurable spaces. A mapping $f : \Omega \mapsto \Lambda$ is said to be *measurable* if the following property holds:

$$f^{-1}(G) \in \mathcal{F} \quad \text{whenever} \quad G \in \mathcal{G}.$$

Definition A.2.2. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A *random variable* is a measurable mapping $\xi : \Omega \mapsto \mathbb{R}$, and a *random vector* is a measurable mapping $\xi : \Omega \mapsto \mathbb{R}^N$.

Given a random vector $\xi : \Omega \mapsto \mathbb{R}^N$ defined on a probability space $(\Omega, \mathcal{F}, P)$, we will be particularly concerned with the composite mapping $\mu = P \circ \xi^{-1}$. Since ξ is measurable we have $\xi^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R}^N)$, and hence it is easily seen that μ is in fact a Borel probability measure on $\mathbb{R}^N$. We will refer to μ as the *induced probability measure* on $\mathbb{R}^N$.

A.3 Expectations

Definition A.3.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\xi : \Omega \mapsto \mathbb{R}^N$ be a random vector. The *expectation* of ξ is

$$\mathbb{E}[\xi] = \int_\Omega \xi(\omega) P(d\omega) \tag{A.3.1}$$

Remark A.3.1. A formal definition of the integral in (A.3.1) is beyond the scope of our presentation here. Let us just note that the classical Riemann integral was extended by Lebesgue to apply to functions on more general spaces and with respect to general measures. For details, we refer to the textbooks by Dudley [43] and Hoffmann-Jørgensen [61].

Typically, stochastic programming models involve functions of random vectors. Given a random vector $\xi : \Omega \mapsto \mathbb{R}^N$ defined on a probability space $(\Omega, \mathcal{F}, P)$ and a measurable function $g : \mathbb{R}^N \mapsto \mathbb{R}$, the composite mapping $g \circ \xi$ is easily seen to be a random variable. Here, the so-called first transformation theorem provides a convenient formula for the expectation of $g \circ \xi$ in terms of the induced probability measure $\mu = P \circ \xi^{-1}$.

Proposition A.3.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\xi : \Omega \mapsto \mathbb{R}^N$ be a random vector, let $\mu = P \circ \xi^{-1}$ be the induced probability measure on $\mathbb{R}^N$, and let $g : \mathbb{R}^N \mapsto \mathbb{R}$ be a measurable function. Then*

$$\mathbb{E}[g(\xi)] = \int_\Omega g(\xi(\omega))P(d\omega) = \int_{\mathbb{R}^N} g(t)\mu(dt) \tag{A.3.2}$$

A.4 Weak Convergence

A recurring subject throughout the thesis is that of stability analysis of stochastic programming models, which concerns certain continuity properties of optimal solutions when the underlying probability measure is subjected to perturbations. When dealing with such a stability analysis we will be interested in continuity properties of the expectation in (A.3.2) when the distribution μ varies in some subset of $\mathcal{P}(\mathbb{R}^N)$. To this end, we will endow the set $\mathcal{P}(\mathbb{R}^N)$ of all Borel probability measures on $\mathbb{R}^N$ with the notion of weak convergence defined as follows.

Definition A.4.1. Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ and let $\{\mu_n\}_{n=1}^\infty$ be some sequence of probability measures in $\mathcal{P}(\mathbb{R}^N)$. If for any bounded continuous function, $g : \mathbb{R}^N \mapsto \mathbb{R}$, we have

$$\int_{\mathbb{R}^N} g(t)\mu_n(dt) \xrightarrow{n\to\infty} \int_{\mathbb{R}^N} g(t)\mu(dt),$$

then the sequence $\{\mu_n\}_{n=1}^\infty$ is said to converge weakly to μ and we write $\mu_n \xrightarrow{w} \mu$.

A.5 Marginal and Conditional Distributions

Let us now consider a random vector $(\xi_1, \xi_2) : \Omega \mapsto \mathbb{R}^{N_1} \times \mathbb{R}^{N_2}$ defined on a probability space $(\Omega, \mathcal{F}, P)$, and let us define again the induced probability measure $\mu = P\circ(\xi_1, \xi_2)^{-1}$. Furthermore, let π_1 and π_2 be the usual projections from $\mathbb{R}^{N_1}\times\mathbb{R}^{N_2}$ to $\mathbb{R}^{N_1}$ and $\mathbb{R}^{N_2}$, respectively. In this case the induced probability measures $\mu_1 = \mu \circ \pi_1^{-1}$ and $\mu_2 = \mu \circ \pi_2^{-1}$ are referred to as the marginal distributions of ξ_1 and ξ_2, respectively. Also, the (regular) conditional distribution of ξ_1 given ξ_2 is a mapping $\mu_1^2 : \mathcal{B}(\mathbb{R}^{N_1}) \times \mathbb{R}^{N_2} \mapsto \mathbb{R}$ with the following properties:

(i) $\mu_1^2(\cdot, t_2)$ is a Borel probability measure on $\mathbb{R}^{N_1}$ for any $t_2 \in \mathbb{R}^{N_2}$;

(ii) $\mu_1^2(B, \cdot)$ is a measurable function on $\mathbb{R}^{N_2}$ for any $B \in \mathcal{B}(\mathbb{R}^{N_1})$;

(iii) for any $B \in \mathcal{B}(\mathbb{R}^{N_1}\times\mathbb{R}^{N_2})$, we have

$$\mu(B) = \int_{\mathbb{R}^{N_2}} \int_{\mathbb{R}^{N_1}} \mathbf{1}_B(t_1, t_2)\, \mu_1^2(dt_1, t_2)\mu_2(dt_2),$$

where $\mathbf{1}_B$ denotes the indicator function of the set B,

$$\mathbf{1}_B(t_1, t_2) = \begin{cases} 1 & \text{if } (t_1, t_2) \in B; \\ 0 & \text{otherwise.} \end{cases}$$

Bibliography

[1] S. Ahmed, M. Tawarmalani, and N.V. Sahinidis (2000). A finite branch and bound algorithm for two-stage stochastic integer programs. Published online at Stochastic Programming E-Print Series, `http://dochost.rz.hu-berlin.de/speps/`.

[2] M. Albareda-Sambola, M.H. van der Vlerk, and E. Fernández (2002). Exact solutions to a class of stochastic generalized assignment problems. SOM Research Report 02A11, University of Groningen. Published online at Stochastic Programming E-Print Series, `http://dochost.rz.hu-berlin.de/speps/`.

[3] R. Andonov, V. Poirriez, and S. Rajopadhye (2000). Unbounded knapsack problem: Dynamic programming revisited. *European Journal of Operational Research* 123, 394–407.

[4] Z. Artstein and R.J-B. Wets (1994). Stability results for stochastic programs and sensors, allowing for discontinuous objective functions. *SIAM Journal on Optimization* 4, 537–550.

[5] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance* 9, 203–228.

[6] A. Balakrishnan, T.L. Magnanti, A. Shulman, and R.T. Wong (1991). Models for planning capacity expansion in local access telecommunication networks. *Annals of Operations Research* 33, 239–284.

[7] A. Balakrishnan, T.L. Magnanti, and R.T. Wong (1989). A dual-ascent procedure for large-scale uncapacitated network design. *Operations Research* 37, 716–740.

[8] A. Balakrishnan, T.L. Magnanti, and R.T. Wong (1995). A decomposition algorithm for local access telecommunications network expansion planning. *Operations Research* 43, 58–76.

[9] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer (1982). *Non-Linear Parametric Optimization*. Akademie Verlag, Berlin.

[10] B. Bank and R. Mandel (1988). *Parametric Integer Optimization*. Akademie Verlag, Berlin.

[11] E.M.L. Beale (1955). On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society, Series B* 17, 173–184.

[12] J.F. Benders (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik* 4, 238–252.

[13] B. Bereanu (1981). Minimum risk criterion in stochastic optimization. *Economic Computation and Economic Cybernetics Studies and Research* 2, 31–39.

[14] C. Berge (1963). *Topological Spaces*. Macmillian, New York.

[15] R.N. Bhattacharya and R. Ranga Rao (1976). *Normal Approximation and Assymptotic Expansions*. John Wiley & Sons, New York.

[16] D. Bienstock, S. Chopra, O. Günlük, and C-Y. Tsai (1998). Minimum cost capacity installation for multicommodity network flows. *Mathematical Programming* 81, 177–200.

[17] D. Bienstock and O. Günlük (1996). Capacitated network design. Polyhedral structure and computation. *INFORMS Journal on Computing* 8, 243–259.

[18] P. Billingsley (1968). *Convergence of Probability Measures*. John Wiley & Sons, New York.

[19] J.R. Birge (1985). Decomposition and partitioning methods for multi-stage stochastic linear programs. *Operations Research* 33, 989–1007.

[20] J.R. Birge and J.H. Dulá (1991). Bounding separable recourse functions with limited distribution information. *Annals of Operations Research* 30, 277–298.

[21] J.R. Birge and F.V. Louveaux (1988). A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research* 34, 384–392.

[22] J.R. Birge and F.V. Louveaux (1997). *Introduction to Stochastic Programming*. Springer-Verlag, New York.

[23] J.R. Birge and R.J-B. Wets (1987). Computing bounds for stochastic programming problems by means of a generalized moment problem. *Mathematics of Operations Research* 12, 149–162.

[24] C.E. Blair and R.G. Jeroslow (1977). The value function of a mixed integer program: I. *Discrete Mathematics* 19, 121–138.

[25] M. Breton and S.E. Hachem (1995). Algorithms for the solution of stochastic dynamic minimax problems. *Computational Optimization and Applications* 4, 317–345.

[26] M. Breton and S.E. Hachem (1995). A scenario aggregation algorithm for the solution of stochastic dynamic minimax problems. *Stochastics and Stochastics Reports* 53, 305–322.

[27] C.C. Carøe (1998). *Decomposition in stochastic integer programming*. PhD thesis, University of Copenhagen.

[28] C.C. Carøe, A. Ruszczyński, and R. Schultz (1997). Unit commitment under uncertainty via two-stage stochastic programming. In C.C. Carøe and D. Pisinger, eds., *Proceedings NOAS'97*, DIKU Rapport 97/23, pp. 21–30. Department of Computer Science, University of Copenhagen.

[29] C.C. Carøe and R. Schultz (1999). Dual decomposition in stochastic integer programming. *Operations Research Letters* 24, 37–45.

[30] C.C. Carøe and J. Tind (1997). A cutting-plane approach to mixed 0-1 stochastic integer programs. *European Journal of Operational Research* 101, 306–316.

[31] C.C. Carøe and J. Tind (1998). L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming* 83, 451–464.

[32] S. Challinor (2000). An introduction to IP networks. *BT Technology Journal* 18.

[33] S-G. Chang and B. Gavish (1993). Telecommunications network topological design and capacity expansion: Formulations and algorithms. *Telecommunication Systems* 1, 99–131.

[34] S-G. Chang and B. Gavish (1995). Lower bounding procedures for multiperiod telecommunications network expansion problems. *Operations Research* 43, 43–57.

[35] A. Charnes and W.W. Cooper (1959). Chance-constrained programming. *Management Science* 5, 73–79.

[36] G. Dahl and M. Stoer (1998). A cutting plane algorithm for multicommodity survivable network design problems. *INFORMS Journal on Computing* 10, 1–11.

[37] G.B. Dantzig (1955). Linear programming under uncertainty. *Management Science* 1, 197–206.

[38] G.B. Dantzig and A. Madansky (1961). On the solution of two-stage linear programs under uncertainty. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, CA.

[39] G.B. Dantzig and P. Wolfe (1960). The decomposition principle for linear programs. *Operations Research* 8, 101–111.

[40] M.A.H. Dempster, E.A. Medova, and R.T. Thompson (1997). A stochastic programming approach to network planning. In V. Ramaswami and P.E. Wirth, eds., *Teletraffic Contributions for the Information Age. Proceedings of the 15th International Teletraffic Congress - ITC 15*, pp. 329–339. Elsevier Science, Amsterdam.

[41] F. Deutsch, W. Pollul, and I. Singer (1973). On set-valued metric projections, Hahn-Banach extension maps, and spherical image maps. *Duke Mathematical Journal* 40, 355–370.

[42] R.M. Dudley (1976). *Probabilities and Metrics*. Lecture Notes Series No. 45. Aarhus Universitet, Aarhus, Denmark.

[43] R.M. Dudley (1989). *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Pacific Grove.

[44] J. Dupačová (1987). The minimax approach to stochastic programming and an illustrative application. *Stochastics* 20, 73–88.

[45] J. Dupačová (1987). Stochastic Programming with incomplete information: A survey of results on postoptimization and sensitivity analysis. *Optimization* 18, 507–532.

[46] J. Dupačová (1990). Stability and sensitivity analysis for stochastic programming. *Annals of Operations Research* 27, 115–142.

[47] E. Engell, A. Märkert, G. Sand, R. Schultz, and C. Schultz (2001). Online scheduling of multiproduct batch plants under uncertainty. In M. Grötschel, S.O. Krumke, and J. Rambau, eds., *Online Optimization of Large Scale Systems*, pp. 649–676. Springer-Verlag, Berlin.

[48] Yu.M. Ermoliev, A. Gaivoronski, and C. Nedeva (1985). Stochastic optimization problems with partially known distribution functions. *SIAM Journal on Control and Optimization* 23, 697–716.

[49] P.C. Gilmore and R.E. Gomory (1966). The theory and computation of knapsack functions. *Operations Research* 14, 1045–1074.

[50] C.R. Givens and R.M. Shortt (1984). A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal* 31, 231–240.

[51] J.E. Graver (1975). On the foundation of linear and integer programming I. *Mathematical Programming* 9, 207–226.

[52] M. Grötschel, C.L. Monma, and M. Stoer (1995). Design of survivable networks. In M.O. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, eds., *Network Models*, volume 7 of *Handbooks in Operations Research and Management Science*, chapter 10, pp. 617–672. North-Holland, Elsevier Science, Amsterdam.

[53] O. Günlük (1999). A branch-and-cut algorithm for capacitated network design. *Mathematical Programming* 86, 17–39.

[54] O. Günlük and Y. Pochet (2001). Mixing mixed-integer inequalities. *Mathematical Programming* 90, 429–457.

[55] R. Hemmecke and R. Schultz (2003). Decomposition of test sets in stochastic integer programming. *Mathematical Programming* 94, 323–341.

[56] J.L. Higle and S. Sen (1991). Statistical verification of optimality conditions. *Annals of Operations Research* 30, 215–240.

[57] J.L. Higle and S. Sen (1991). Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research* 16, 650–669.

[58] J.L. Higle and S. Sen (1994). Finite master programs in regularized stochastic decomposition. *Mathematical Programming* 67, 143–168.

[59] J.L. Higle and S. Sen (1996). *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*. Kluwer Academic Publishers, Dordrecht, Netherlands.

[60] J.L. Higle and S. Sen (1999). Statistical approximations for stochastic linear programming problems. *Annals of Operations Research* 85, 173–192.

[61] J. Hoffmann-Jørgensen (1994). *Probability with a View Toward Statistics*, volume I. Chapman & Hall, New York.

[62] K. Holmberg and D. Yuan (2001). Optimization of internet protocol network design and routing. Research Report LiTH-MAT-R-2001-07, Linköping Institute of Technology, Department of Mathematics.

[63] M. Iri (1971). On an extension of the max-flow min-cut theorem to multicommodity flows. *Journal of the Operations Research Society of Japan* 13, 129–135.

[64] K.O. Jörnsten, M. Näsberg, and P.A. Smeds (1985). Variable splitting - a new Lagrangean relaxation approach to some mathematical programming models. Research Report LiTH-MAT-R-85-04, Linköbing Institute of Technology, Department of Mathematics.

[65] P. Kall (1987). On approximations and stability in stochastic programming. In J. Guddat, H.Th. Jongen, B. Kummer, and F. Nožička, eds., *Parametric Optimization and Related Topics*, pp. 387–407. Akademie Verlag, Berlin.

[66] P. Kall (1991). An upper bound for stochastic linear programming using first and total second moments. *Annals of Operations Research* 30, 267–276.

[67] P. Kall and S.W. Wallace (1994). *Stochastic Programming*. John Wiley & Sons, Chichester, UK.

[68] A.F. Karr (1983). Extreme points of certain sets of probability measures, with applications. *Mathematics of Operations Research* 8, 74–85.

[69] J.H.B. Kempermann (1968). The general moment problem, a geometric approach. *Annals of Mathematical Statistics* 39, 93–122.

[70] A.I. Kibzun and Y.S. Kan (1996). *Stochastic Programming Problems with Probability and Quantile Functions*. John Wiley & Sons, Chichester, UK.

[71] K.C. Kiwiel (1985). *Methods of Descent for Nondifferentiable Optimization*. Number 1133 in Lecture Notes in Mathematics. Springer-Verlag, Berlin.

[72] D. Klatte (1985). On the stability of local and global optimal solutions in parametric problems of nonlinear programming. Part I and Part II. Seminarbericht Nr. 75, Humboldt-Universität zu Berlin, Sektion Mathematik.

[73] D. Klatte (1987). A note on quantitative stability results in nonlinear optimization. In K. Lommatzsch, ed., *Proceedings of the 19. Jahrestagung Mathematische Optimierung*, Seminarbericht Nr. 90, pp. 77–86. Humboldt-Universität zu Berlin, Sektion Mathematik.

[74] W.K. Klein Haneveld, L. Stougie, and M.H. van der Vlerk (1995). On the convex hull of the simple integer recourse objective function. *Annals of Operations Research* 56, 209–224.

[75] W.K. Klein Haneveld, L. Stougie, and M.H. van der Vlerk (1996). An algorithm for the construction of convex hulls in simple integer recourse programming. *Annals of Operations Research* 64, 67–81.

[76] W.K. Klein Haneveld, L. Stougie, and M.H. van der Vlerk (1997). Convex approximations for simple integer recourse models by perturbing the underlying distribution. SOM Research Report 97A19, University of Groningen.

[77] W.K. Klein Haneveld, L. Stougie, and M.H. van der Vlerk (1997). Convex simple integer recourse models. SOM Research Report 97A10, University of Groningen.

[78] W.K. Klein Haneveld and M.H. van der Vlerk (1994). On the expected value function of a simple integer recourse problem with random technology matrix. *Journal of Computational and Applied Mathematics* 56, 45–53.

[79] W.K. Klein Haneveld and M.H. van der Vlerk (1999). Stochastic integer programming: General models and algorithms. *Annals of Operations Research* 85, 39–57.

[80] K. Kuratowski (1966). *Topology*. Academic Press, New York.

[81] M. Laguna (1998). Applying robust optimization to capacity expansion of one location in telecommunications with demand uncertainty. *Management Science* 44, S101–S110.

[82] G. Laporte and F.V. Louveaux (1993). The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters* 13, 133–142.

[83] G. Laporte, F.V. Louveaux, and H. Mercure (1992). The vehicle routing problem with stochastic travel times. *Transportation Science* 26, 161–170.

[84] G. Laporte, F.V. Louveaux, and H. Mercure (1994). A priori optimization of the probabilistic traveling salesman problem. *Operations Research* 42, 543–549.

[85] G. Laporte, F.V. Louveaux, and H.L. van Hamme (1994). Exact solution of a stochastic location problem by an integer L-shaped algorithm. *Transportation Science* 28, 95–103.

[86] F.V. Louveaux and M.H. van der Vlerk (1993). Stochastic programming with simple integer recourse. *Mathematical Programming* 61, 301–3250.

[87] A. Løkketangen and D.L. Woodruff (1996). Progressive hedging and tabu search applied to mixed integer (0,1) multistage stochastic programming. *Journal of Heuristics* 2, 111–128.

[88] T.L. Magnanti, P. Mirchandani, and R. Vachani (1993). The convex hull of two core capacitated network design problems. *Mathematical Programming* 60, 233–250.

[89] T.L. Magnanti, P. Mirchandani, and R. Vachani (1995). Modeling and solving the two-facility capacitated network loading problem. *Operations Research* 43, 142–157.

[90] E.A. Medova (1998). Chance-constrained stochastic programming for integrated services network management. *Annals of Operations Research* 81, 213–229.

[91] A. Merchant and B. Sengupta (1995). Assignment of cells to switches in PCS networks. *IEEE/ACM Transactions on Networking* 3, 521–526.

[92] M. Minoux (1981). Optimum synthesis of a network with non-simultaneous multicommodity flow requirements. In P. Hansen, ed., *Studies on Graphs and Discrete Programming*, volume 11 of *Annals of Discrete Mathematics*, pp. 269–277. North-Holland Publishing Company, Amsterdam.

[93] P. Mirchandani (2000). Projections of the capacitated network loading problem. *European Journal of Operational Research* 122, 534–560.

[94] S.A. MirHassani, C. Lucas, G. Mitra, E. Messina, and C.A. Poojari (2000). Computational solution of capacity planning models under uncertainty. *Parallel Computing* 26, 511–538.

[95] J.M. Mulvey and A. Ruszczyński (1995). A new scenario decomposition method for large-scale stochastic optimization. *Operations Research* 43, 477–490.

[96] J.M. Mulvey, R.J. Vanderbei, and S.J. Zenios (1995). Robust optimization of large-scale systems. *Operations Research* 43, 264–281.

[97] G.L. Nemhauser and L.A. Wolsey (1988). *Integer and Combinatorial Optimization*. Wiley-Interscience, New York.

[98] V.I. Norkin, Yu.M. Ermoliev, and A. Ruszczyński (1998). On optimal allocation of indivisibles under uncertainty. *Operations Research* 46, 381–395.

[99] V.I. Norkin, G.Ch. Pflug, and A. Ruszczyński (1998). A branch and bound method for stochastic global optimization. *Mathematical Programming* 83, 425–450.

[100] W. Ogryczak and A. Ruszczyński (1999). From stochastic dominance to mean-risk models: Semideviations as risk measures. *European Journal of Operational Research* 116, 33–50.

[101] W. Ogryczak and A. Ruszczyński (2001). On consistency of stochastic dominance and mean-semideviation models. *Mathematical Programming* 89, 217–232.

[102] W. Ogryczak and A. Ruszczyński (2002). Dual stochastic dominance and quantile risk measures. *International Transactions in Operational Research* 9, 661–680.

[103] W. Ogryczak and A. Ruszczyński (2002). Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization* 13, 60–78.

[104] K. Onaga and O. Kakusho (1971). On feasibility conditions of multicommodity flows in networks. *IEEE Transactions in Circuit Theory CT-18* 4, 425–429.

[105] M. Padberg and G. Rinaldi (1991). A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems. *SIAM Review* 33, 60–100.

[106] M.V.F. Pereira and L.M.V.G. Pinto (1985). Stochastic optimization of a multireservoir hydroelectric system - A decomposition approach. *Water Resources Research* 21, 779–792.

[107] A. Prékopa (1995). *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, Netherlands.

[108] S.T. Rachev (1991). *Probability Metrics and the Stability of Stochastic Models*. John Wiley & Sons, New York.

[109] S.T. Rachev and W. Römisch (2002). Quantitative stability in stochastic programming: The method of probability metrics. *Mathematics of Operations Research* 27, 792–818.

[110] E. Raik (1971). Qualitative research into the stochastic nonlinear programming problems. *Eesti NSV Teaduste Akademia Toimetised (News of the Estonian Academy of Sciences) Füüs Mat.* 20, 8–14. In Russian.

[111] E. Raik (1972). On the stochastic programming problem with the probability and quantile functionals. *Eesti NSV Teaduste Akademia Toimetised (News of the Estonian Academy of Sciences) Füüs Mat.* 21, 142–148. In Russian.

[112] M. Riis and K.A. Andersen. Multiperiod capacity expansion of a telecommunications connection with uncertain demand. *Computers & Operations Research* (to appear). Available at `http://www.sciencedirect.com`.

[113] M. Riis and K.A. Andersen (2002). Applying the minimax criterion in stochastic recourse programs. Working Paper 2002/4, University of Aarhus, Department of Operations Research. Published online at Stochastic Programming E-Print Series, `http://dochost.rz.hu-berlin.de/speps/`.

[114] M. Riis and K.A. Andersen (2002). Capacitated network design with uncertain demand. *INFORMS Journal on Computing* 14, 247–260.

[115] M. Riis and J. Lodahl (2002). A bicriteria stochastic programming model for capacity expansion in telecommunications. *Mathematical Methods of Operations Research* 56, 83–100.

[116] M. Riis and R. Schultz (2003). Applying the minimum risk criterion in stochastic recourse programs. *Computational Optimization and Applications* 24, 267–287.

[117] M. Riis, A.J.V. Skriver, and J. Lodahl (2001). Network planning in telecommunications: A stochastic programming approach. Working Paper 2001/2, University of Aarhus, Department of Operations Research. Available at `http://home.imf.au.dk/riis/`.

[118] M. Riis, A.J.V. Skriver, and S.F. Møller (2003). Internet protocol network design with uncertain demand. Working Paper 2003/2, University of Aarhus, Department of Operations Research. Available at `http://home.imf.au.dk/riis/`.

[119] A.H.G. Rinnooy Kan and L. Stougie (1988). Stochastic integer programming. In Yu.M. Ermoliev and R.J-B. Wets, eds., *Numerical Techniques for Stochastic Optimization*, pp. 201–213. Springer-Verlag, Berlin.

[120] M. Rios, V. Marianov, and M. Gutierrez (2000). Survivable capacitated network design problem: new formulation and Lagrangean relaxation. *Journal of the Operational Research Society* 51, 574–582.

[121] S.M. Robinson (1987). Local epi-continuity and local optimization. *Mathematical Programming* 37, 208–222.

[122] S.M. Robinson and R.J-B. Wets (1987). Stability in two-stage stochastic programming. *SIAM Journal on Control and Optimization* 25, 1409–1416.

[123] R.T. Rockafellar (1970). *Convex Analysis*. Princeton University Press, Princeton, NJ.

[124] R.T. Rockafellar and R.J-B. Wets (1991). Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research* 16, 119–147.

[125] W. Römisch and R. Schultz (1991). Distribution sensitivity in stochastic programming. *Mathematical Programming* 50, 197–226.

[126] W. Römisch and R. Schultz (1991). Stability analysis for stochastic programs. *Annals of Operations Research* 30, 241–266.

[127] W. Römisch and R. Schultz (1993). Stability of solutions for stochastic programs with complete recourse. *Mathematics of Operations Research* 18, 590–609.

[128] W. Römisch and R. Schultz (1996). Lipschitz stability for stochastic programs with complete recourse. *SIAM Journal on Optimization* 6, 531–547.

[129] W. Römisch and R. Schultz (2001). Multistage stochastic integer programs: An introduction. In M. Grötschel, S.O. Krumke, and J. Rambau, eds., *Online Optimization of Large Scale Systems*, pp. 581–600. Springer-Verlag, Berlin.

[130] W. Römisch and A. Wakolbinger (1987). Obtaining convergence rates for approximations in stochastic programming. In J. Guddat, H.Th. Jongen, B. Kummer, and F. Nožička, eds., *Parametric Optimization and Related Topics*, pp. 327–343. Akademie Verlag, Berlin.

[131] C.H. Rosa and A. Ruszczyński (1996). On augmented Lagrangian decomposition methods for multistage stochastic programs. *Annals of Operations Research* 64, 289–309.

[132] A. Ruszczyński (1986). A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming* 35, 309–333.

[133] A. Ruszczyński (1999). Some advances in decomposition methods for stochastic linear programming. *Annals of Operations Research* 85, 153–172.

[134] A. Ruszczyński and A. Swietanowski (1997). Accelerating the regularized decomposition method for two stage stochastic linear problems. *European Journal of Operational Research* 101, 328–342.

[135] D. Saha, A. Mukherjee, and P.S. Bhattacharya (2000). A simple heuristic for assignment of cells to switches in a PCS network. *Wireless Personal Communications* 12, 209–224.

[136] I. Saniee (1995). An efficient algorithm for the multiperiod capacity expansion of one location in telecommunications. *Operations Research* 43, 187–190.

[137] R. Schultz. Probability objectives in stochastic programs with recourse. To appear in *Proceedings of the 20th IFIP-TC7 Conference on System Modelling and Optimization (Trier, Germany, July 2001).* Kluwer Academic Publishers.

[138] R. Schultz (1992). Continuity and stability in two-stage stochastic integer programming. In K. Marti, ed., *Stochastic Optimization, Numerical Methods and Technical Applications*, volume 379 of *Lecture Notes in Economics and Mathematical Systems*, pp. 81–92. Springer-Verlag, Berlin.

[139] R. Schultz (1993). Continuity properties of expectation functions in stochastic integer programming. *Mathematics of Operations Research* 18, 578–589.

[140] R. Schultz (1995). On structure and stability in stochastic programs with random technology matrix and complete integer recourse. *Mathematical Programming* 70, 73–89.

[141] R. Schultz (1996). Rates of convergence in stochastic programs with complete integer recourse. *SIAM Journal on Optimization* 6, 1138–1152.

[142] R. Schultz (2000). Some aspects of stability in stochastic programming. *Annals of Operations Research* 100, 54–84.

[143] R. Schultz, L. Stougie, and M.H. van der Vlerk (1998). Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis reductions. *Mathematical Programming* 83, 229–252.

[144] R. Schultz and S. Tiedemann (2002). Risk aversion via excess probabilities in stochastic programs with mixed-integer recourse. Preprint 531-2002, Gerhard-Mercator-Universität Duisburg, Institut für Mathematik. Published at Stochastic Programming E-Print Series, `http://dochost.rz.hu-berlin.de/speps/`.

[145] S. Sen, R.D. Doverspike, and S. Cosares (1994). Network planning with random demand. *Telecommunication Systems* 3, 11–30.

[146] A. Shapiro (1994). Quantitative stability in stochastic programming. *Mathematical Programming* 67, 99–108.

[147] A. Shapiro and A. Kleywegt (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software* 17, 523–542.

[148] R.E. Steuer (1986). *Multiple Criteria Optimization: Theory, Computation and Application*. John Wiley & Sons, New York.

[149] M. Stoer and G. Dahl (1994). A polyhedral approach to multicommodity survivable network design. *Numerische Mathematik* 68, 149–167.

[150] L. Stougie (1987). *Design and analysis of algorithms for stochastic integer programming*. CWI Tract 37. Center for Mathematics and Computer Science, Amsterdam.

[151] L. Stougie and M.H. van der Vlerk (1997). Stochastic integer programming. In M. Dell'Amico, F. Maffioli, and S. Martello, eds., *Annotated Bibliographies in Combinatorial Optimization*, chapter 9, pp. 127–141. John Wiley & Sons, New York.

[152] S. Takriti and S. Ahmed (2002). Managing short-term electricity contracts under uncertainty: A minimax approach. Available at `http://www.isye.gatech.edu/~sahmed`.

[153] S. Takriti, J.R. Birge, and E. Long (1996). A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems* 11, 1497–1508.

[154] S. Tiedemann (2001). *Probability functionals and risk aversion in stochastic integer programming*. Master's thesis, Gerhard-Mercator-Universität, Duisburg.

[155] E.C. Tzifa, V.P. Demestichas, M.E. Theologou, and M.E. Anagnostou (1999). Design of the access network segment of future mobile communications systems. *Wireless Personal Communications* 11, 247–268.

[156] M.H. van der Vlerk (1995). *Stochastic programming with integer recourse*. PhD thesis, Rijksuniversiteit Groningen.

[157] M.H. van der Vlerk (1996-2003). Stochastic Programming Bibliography. Available at `http://mally.eco.rug.nl/spbib.html`.

[158] M.H. van der Vlerk (2002). Convex approximations for complete integer recourse models. SOM Research Report 02A21, University of Groningen. Published online at Stochastic Programming E-Print Series, `http://dochost.rz.hu-berlin.de/speps/`.

[159] R.M. Van Slyke and R.J-B. Wets (1969). L-shaped linear programs with applications to optimal control and stochastic linear programming. *SIAM Journal of Applied Mathematics* 17, 638–663.

[160] D. Walkup and R.J-B. Wets (1967). Stochastic programs with recourse. *SIAM Journal on Applied Mathematics* 15, 1299–1314.

[161] S.W. Wallace (2000). Decision making under uncertainty: Is sensitivity analysis of any use? *Operations Research* 48, 20–25.

[162] J. Wang (1985). Distribution sensitivity analysis for stochastic programs with complete recourse. *Mathematical Programming* 31, 286–297.

[163] R.J-B. Wets (1974). Stochastic programs with fixed recourse: the equivalent deterministic program. *SIAM Review* 16, 309–339.

[164] R.J-B. Wets (1983). Solving stochastic programs with simple recourse. *Stochastics* 10, 219–242.

[165] G.A. Whitmore and M.C. Findlay, eds. (1978). *Stochastic Dominance: An Approach to Decision-Making Under Risk*. D.C. Heath and Company, Lexington, MA.

[166] R.D. Wollmer (1980). Two stage linear programming under uncertainty with 0-1 integer first stage variables. *Mathematical Programming* 19, 279–288.

[167] J. Žáčková (1966). On minimax solutions of stochastic linear programming problems. *Časopsis pro Pěstování Matematiky* 91, 423–430.

[168] W.T. Ziemba (1974). Stochastic programs with simple recourse. In P.L. Hammer and G. Zoutendijk, eds., *Mathematical Programming in Theory and Practice*, pp. 213–273. North-Holland Publishing Company, Amsterdam.