International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

**Interim Report IR-07-041**

# Modeling of Competition and Collaboration Networks under Uncertainty: Stochastic Programs with Resource and Bilevel Structure

*Alexei Gaivoronski (Alexei.Gaivoronski@iot.ntnu.no)*
*Adrian Werner (Adrian.Werner@ucd.ie)*

**Approved by**

*Marek Makowski (marek@iiasa.ac.at)*
Leader, Integrated Modeling Environment Project

December 2007

# Foreword

The performance of communication and other networks may be significantly (intentionally or unintentionally) affected (or even endangered) by a second party. This raises critical issue about endogenous uncertainty: besides traditional exogenous (environmental) uncertainty the key issue is a proper treatment of the uncertainty which may potentially be created by a second party.

This paper is devoted to modeling such type of uncertainty by using the "leader-follower" terminology. It was shown that evaluations of potential responses by a follower can be formalized by the so-called stochastic bilevel programming models. Resulting new type of stochastic optimization models typically have the nonconvex and even discontinuous character. The paper develops a promising approach relying on a partitioning strategy to cope with nonconvexities and stochastic quasigradient methods to cope with multidimensional often not perfectly known probability distributions, potential discontinuities and implicit dependencies. It should be stressed that although the paper focuses mainly on methods and algorithms for this new particular class of problems, the proposed approach opens-up possibilities for effectively solving practical network problems that cannot be solved by existing methods.

The results reported in this paper were intensively discussed and formalized during the visit of Alexei Gaivoronski to the IME Project at IIASA. These results provide a good basis for further research, which is in particular relevant to analysis of network robustness within the forthcoming IIASA initiative on the Fragility of Critical Infrastructures.

## Abstract

We analyze stochastic programming problems with recourse characterized by a bilevel structure. Part of the uncertainty in such problems is due to actions of other actors such that the considered decision maker needs to develop a model to estimate their response to his decisions. Often, the resulting model exhibits connecting constraints in the leader's (upper-level) subproblem. It is shown that this problem can be formulated as a new class of stochastic programming problems with equilibrium constraints (SMPEC). Sufficient optimality conditions are stated. A solution algorithm utilizing a stochastic quasi-gradient method is proposed, and its applicability extensively explained by practical numerical examples.

**Key-words:** Stochastic mathematical program with equilibrium constraints, decision making under uncertainty, bilevel structure, Stochastic Quasigradient Method.

## Acknowledgments

The research presented in this paper was partially supported by Science Foundation Ireland under Grant No. 03/CE3/I405 as part of the Centre for Telecommunications Value Chain Research (CTVR). Alexei Gaivoronski appreciates the hospitality and fruitful discussions with colleagues from IME Project during his visit to IIASA.

## About the Authors

Alexei Gaivoronski is a Professor at Department of Industrial Economics and Technology Management, Norwegian University, Norway. His scientific interests are focused on mathematical methodologies for optimal decisions under risk and uncertainty, spanning the relevant aspects of mathematics, optimization, finance, management science, economics. The main focus of his work is modeling, analysis and optimization of uncertain and stochastic systems and, in particular, stochastic programming.

Alexei has published more than 60 papers on these topics in international journals and book chapters. He collaborates with industry (telecom, finance, energy) both nationally in Norway and internationally in Europe.

Adrian Werner got his Ph.D. in Operations Research in 2005 from the Norwegian University of Science and Technology. His research interests include stochastic optimization, network and telecommunications models. After getting his Ph.D. he has hold a postdoctoral position at the Center for Telecommunications Value Chain Research, University College Dublin. Recently he has joined the Research Center SINTEF, Norway.

# Contents

# List of Tables

# List of Figures

# Modeling of Competition and Collaboration Networks under Uncertainty: Stochastic Programs with Resource and Bilevel Structure

*Alexei Gaivoronski (Alexei.Gaivoronski@iot.ntnu.no)*[*]
*Adrian Werner (Adrian.Werner@ucd.ie)*[** ***]

## 1 Introduction

In this paper, we consider stochastic programming problems where the uncertainty which the decision maker faces can be further classified into two main categories. The first category is the uncertainty traditionally found in stochastic programming and can be described by random parameters with known or unknown probability distributions, we call it environmental uncertainty. The second category describes the uncertainty created by actions of other decision makers. In order to cope with this type of uncertainty, the first decision maker, the leader, develops a model which describes the decision process of the other actors, called followers. This allows the leader to predict the follower responses to his choices. One can expect that such a model will never be fully precise and in its turn will include uncertain parameters traditional to stochastic programming. Our main objective is to show how stochastic programming concepts can be utilized and generalized for the treatment of such problems. In this paper we do the first steps in pursuing this research program. As a starting point, we concentrate on the case when the leader evaluates (models) the actions of a single follower. We extend the concept of stochastic programming problems with recourse to this case and study properties of the resulting problems. Furthermore, we develop a solution algorithm which combines the stochastic quasi-gradient method with a Lagrangian approach.

Deterministic bilevel programming problems and their generalizations, mathematical programs with equilibrium constraints (MPECs) were studied intensely during the past decades [Dem02, LPR96] and a variety of solution methods has been developed [FL04, FLRS02, JR03, KO04, LS04, ZL01]. The bilevel structure of the problems complicates their analysis. Taking into account the follower's response, the leader's objective function is generally neither convex nor differentiable. If the leader's constraints also include the follower's response (so called connecting upper-level constraints) then the region of feasible leader decisions may not even be connected. In this paper we will explicitly take into account such connecting upper-level constraints. Our viewpoint is motivated

[*]Integrated Modeling Environment Project of IIASA.
[**]Norwegian University of Science and Technology, Trondheim, Norway.
[***]Centre for Telecommunications Value Chain Research and National Institute for Technology Management, University College Dublin, Carysfort Avenue, Blackrock, Co. Dublin, Ireland.

by a number of applications, for example in telecommunications [AGW06], in energy and power management [GR02] or more generally in agency theory [Mir99, WG05]. We suggest a partitioning of the leader's feasible region into convex segments and the application of a gradient algorithm restricted on such segments. This way we can deal with the mentioned complications.

A natural extension of the deterministic models allows for the inclusion of uncertain model parameters, resulting in stochastic bilevel programming problems [Wyn01] and, more generally, stochastic mathematical programs with equilibrium constraints (SMPEC). Due to the complex model structure, the environmental uncertainty may enter the problem at various points and the relations between the subproblems may be designed in a number of ways. This leads to several, quite distinct, types of SMPECs discussed in the literature. Main classes recently investigated comprise a so-called lower-level wait-and-see structure and here-and-now models. In the first type, the upper-level or leader's decisions must be found before the uncertainty reveals while the lower-level or followers' response is made after observing the environmental uncertainty. This approach has been considered in work by Shapiro and Xu [Sha06, SX05, Xu06]. The second class is comparable to the common approach known from stochastic programming, all decisions must be made before the environmental uncertainty can be observed. Some special versions of that class were studied by Lin et al [LCFar, LF06] while another formulation is the subject of research by Birbil et al [BGL04a, BGL04b]. Lin and Fukushima [LF06] consider an interesting variation of this class by introducing a recourse variable which may correct a violation of the complementarity constraint. In contrast, the models analyzed in this paper combine here-and-now and wait-and-see features in the sense of separate stochastic programming problems with recourse in the leader and follower subproblems. Hence, we describe yet another class which, to our knowledge, has not been studied until now.

The concept of stochastic programming problems with recourse [EW88b, Wet89] enables us to take account for dynamic aspects. Patriksson and Wynter [PW99] showed that both two-stage stochastic programming problems and bilevel programming problems are basically similar subclasses of (S)MPEC and can be reformulated as such. However, this does not apply to the class of models considered here, especially when the bilevel structure is present at both stages, possibly even with further interrelations between decision variables of the single subproblems. The first- and second-stage equilibrium problems need to be treated separately due to the nonanticipativity property. We study several problem variations arising from the leader-follower interaction and show that they can be reduced to one common formulation of a two-stage stochastic programming problem with recourse and a complementarity constraint. Further examples of this class of SMPEC, partially with nonlinear constraints, are discussed in [WW07].

Suggestions for SMPEC solution approaches comprise smoothing or penalty methods [EP04, LCFar] or the utilization of a finite number of scenarios and deterministic equivalent formulations [PW99]. This results in large deterministic problems which are computationally expensive for problems of a realistic size. Another type of solution methods employs approximations by deterministic equivalents obtained by sampling methods [BGL04b, LCFar, Sha06, SX05]. In contrast, our approach focuses on the stochastic programming features of the problem. This way it is possible to apply the stochastic programming methodology [EW88a, SR03] directly to the two-stage problem. In particular, we employ techniques using sampling during the solution process, such as stochastic quasi-gradient methods [Erm88, Gai88, Gai04]. This accommodates various representations of the uncertain variables, for example continuous distributions. Also problems where a calculation of deterministic equivalents is difficult, for example due to multidimensional, complicated or even not perfectly known distributions, or certain types of discontinuous problems can be considered. Therefore, our viewpoint enables a more comprehensive treatment of the uncertainty and more complex problem structures.

Finally, it should be noted that so far only few authors (Lin et al [LCFar] and Shapiro and Xu [SX05]) reported on actual results of numerical experiments and on experience with SMPEC solution approaches. Moreover, the studied examples were typically quite small in size. A general comparison of the performance of different approaches would require similar types of the underlying test problems which, however, is not the case.

The following section defines notations and reviews deterministic concepts which form the basis for our further discussion. Section 3 studies two-stage stochastic programming problems with a bilevel structure and different degrees of complexity. Sufficient optimality conditions are stated and a solution algorithm is developed utilizing a stochastic quasi-gradient method. A numerical illustration of the approach is given in Section 4. Section 5 concludes the paper.

# 2 Notations and deterministic concepts

In this section, we introduce some notations and deterministic concepts which are necessary for the subsequent analysis of the stochastic programming problems. We start with a general problem formulation. By refining the assumptions on the problem functions we proceed then to the problem type studied in the remainder of this paper.

Consider the following deterministic optimization problem

$$\begin{aligned} &\min_{y \in Y} F(y, z) \\ &G(y, z) \leq 0 \end{aligned} \tag{1}$$

where the considered decision maker directly controls the variables $y \in Y \subseteq R^n$. The variables $z \in Z \subseteq R^m$ denote the response of another decision maker to these decisions $y$ and are determined by the parametric optimization problem

$$\begin{aligned} &\min_{z \in Z} f(y, z) \\ &g(y, z) \leq 0 \end{aligned} \tag{2}$$

with the parameter $y$. This represents a bilevel programming problem with the upper-level problem (1) and the lower-level problem (2). We assume $F, f : R^n \times R^m \to R^1$, $G : R^n \times R^m \to R^p$ and $g : R^n \times R^m \to R^q$. Furthermore, we assume that the sets $Y$ and $Z$ are convex and compact.

**Assumption 1** *The objective functions $F(y, z)$ and $f(y, z)$ are convex in $y$ and $z$ and at least twice continuously differentiable.*
*The upper-level constraints $G_i(y, z), i = 1, ..., p$, are convex in $y$ and $z$ and at least $C^1$.*
*The lower-level constraints $g_j(y, z), j = 1, ..., q$, are linear in $y$ and $z$.*

The inducible region denotes the set over which the leader may optimize

$$IR = \{y \in Y | \exists z^* \in M(y) : G(y, z^*) \leq 0\}$$

with the lower-level solution set $M(y)$ defined for a given upper-level decision $y^0 \in Y$ by

$$M(y^0) = \arg\min_{z \in Z}\{f(y^0, z) | g(y^0, z) \leq 0\}$$

For a given upper-level parameter $y^0$, we denote the Lagrangian function of the lower-level problem (2) by

$$L(y^0, z^0, \lambda^0, \mu^0) = f(y^0, z^0) + (\lambda^0)^T g(y^0, z^0)$$

with $z^0 \in M(y^0)$ and the Lagrange multipliers $\lambda^0$. Furthermore, we define the following index sets:

$$\begin{aligned}
I_C &= I_C(y^0) = \{i \in \{1, ..., q\} | g_i(y^0, z^0) = 0\} \\
I_L &= I_L(y^0) = \{i \in \{1, ..., q\} | \lambda_i^0 = 0\} \\
\overline{I}_C &= \{1, ..., q\} \setminus I_C \\
\overline{I}_L &= \{1, ..., q\} \setminus I_L
\end{aligned}$$

Problem (1) – (2) exhibits several features prohibiting a direct application of, for example, gradient solution methods. The first feature is that the leader's objective function depends also on the response of the follower. Even if $F(y, z)$ is convex and differentiable with respect to both $y$ and $z$, the function $F(y, z(y))$ may be nondifferentiable and nonconvex in $y$. The second important feature is the presence of connecting upper-level constraints. Their feasibility can be investigated only after the follower's response has been determined. Under certain assumptions, the lower-level solution function $z(y)$ is continuous. However, there may exist responses $z(y)$ which do not satisfy the upper-level constraints $G(y, z) \leq 0$. A consequence is that the inducible region may be not convex, even not connected. Then the convergence of the solution algorithm can not be guaranteed.

**Definition 1** *Consider the problem*

$$\begin{aligned}
&\min_x f(x) \\
&g(x) \leq 0 \\
&h(x) = 0
\end{aligned}$$

*This problem satisfies the Slater constraint qualification if there exists a point $x^0$ such that $g(x^0) < 0$ and $h(x^0)$ is affine.*

**Assumption 2** *The lower-level problem (2) satisfies the Slater constraint qualification for any given feasible upper-level decision $y^0$. Furthermore, the optimal solution $z^0$ of this problem is unique.*

With this assumption the follower's response on a given upper-level decision can be expressed using the Karush Kuhn Tucker optimality conditions on problem (2) and substituted in the upper-level problem (1)

$$\min_{y,z,\lambda} F(y,z) \tag{3a}$$
$$G(y,z) \leq 0$$
$$\nabla_z f(y,z) + \lambda^T \nabla_z g(y,z) = 0$$
$$\lambda^T g(y,z) = 0 \tag{3b}$$
$$g(y,z) \leq 0 \tag{3c}$$
$$\lambda \geq 0 \tag{3d}$$

This one-level nonlinear programming problem represents a Mathematical Programming Problem with Equilibrium Constraints (MPEC). It is ill-posed due to the equilibrium or complementarity constraint (3b). There exists no feasible solution which strictly satisfies all inequalities. Therefore the usual constraint qualifications from nonlinear programming such as the Mangasarian-Fromowitz Constraint Qualification are violated at every feasible point [CF95]. An approach to deal with this difficulty is to reformulate problem (3) by replacing the complementarity constraint (3b), together with constraints (3c) and (3d), by

$$\min\{-g(y^0, z^0), \lambda^0\} = 0 \tag{4}$$

where the minimum is taken componentwise. Given a feasible point $y^0$, the index sets $I_C = I_C(y^0)$ and $I_L = I_L(y^0)$ are defined. Then the nonsmooth constraint (4) can be substituted, for example, by the smooth constraints

$$\begin{aligned} &g_i(y,z) = 0, \qquad i \in I_C \\ &g_i(y,z) \leq 0, \qquad I_L \cap \overline{I}_C \\ &\lambda_i = 0, \qquad i \in I_L \\ &\lambda_i \geq 0, \qquad i \in \overline{I}_L \cap I_C \end{aligned}$$

This way, an ordinary nonlinear programming problem is obtained, the Tightened Nonlinear Program (TNLP) [Fle05, SS00]. It describes a subset of the feasible set of the one-level problem (3) and thus of the original problem (1) – (2). A local optimal solution $x^* = (y^*, z^*, \lambda^*)$ of (1) – (2) is also locally optimal for the TNLP with the according index sets $I_C(y^*)$ and $I_L(y^*)$.
These considerations suggest a partitioning of the inducible region into segments comprising all upper-level decisions $y$ with the same characteristic of the response $z(y)$, i.e. with the same indices of active lower-level constraints and of zero Lagrange multipliers. The partitioning strategy results in a finite number of segments and motivated the development of the solution algorithm in the subsequent section.

**Definition 2** *A segment $Y^s$ is defined by*

$$\begin{aligned} &Y^s = \{y \in Y | I_C^s(y) = I_1^s, I_L^s(y) = I_2^s\} \\ &I_C^s(y) = \{i \in \{1,...,q\} | g_i(y, z(y)) = 0\} \\ &I_L^s(y) = \{i \in \{1,...,q\} | \lambda_i(y) = 0\} \\ &I_1^s, I_2^s \in 2^{\{1,...,q\}} \end{aligned}$$

*where $2^{\{1,\dots,q\}}$ denotes the family of all subsets of the index set $\{1, ..., q\}$.*

The union of all such segments is the upper-level domain $Y$. However, the inducible region

$$IR = \bigcup_s \{y \in Y^s | G(y, z(y)) \le 0\} = \{y \in Y | G(y, z(y)) \le 0\}$$

may be disconnected. This is demonstrated in the following example.

**Example 3** *Consider the problem*

$$\min_y y + z^* \tag{5a}$$

$$z^* \ge 2 \tag{5b}$$

$$y \ge 0 \tag{5c}$$

$$z^* \in \arg\min_{z \in \boldsymbol{R}} \{z | y + z \ge 3,\ y - z \le 3,\ z \ge 1\} \tag{5d}$$

*The optimal solution of the lower-level problem* (5d) *is*

$$z(y) = \begin{cases} y - 3, & 4 \le y \\ 1, & 2 \le y \le 4 \\ -y + 3, & y \le 2 \end{cases}$$

*However, only for $y \in [0, 1] \cup [5, \infty)$ the upper-level constraint* (5b) *is satisfied.*

In order to apply a gradient algorithm we need some properties of the segments.

**Proposition 4** *Assume that the following conditions are satisfied:*

1. *Assumptions 1 and 2 hold,*

2. *for $y \in$ ri $Y^s$ and the response $z$ the Karush Kuhn Tucker conditions on the lower-level problem* (2)

$$\begin{aligned} &\nabla_z f(y, z) + \lambda^T \nabla_z g(y, z) = 0 \\ &\lambda^T g(y, z) = 0 \\ &g(y, z) \le 0 \\ &\lambda \ge 0 \end{aligned} \tag{6}$$

   *are satisfied with strict complementarity.*

*Then the upper-level objective function $F(y, z(y))$ is continuously differentiable on the relative interior ri $Y^s$ of the segment.*

**Proof.** Consider an upper-level decision $y^0 \in$ ri $Y^s$. With Assumption 2 the response $z = z(y)$ is uniquely determined for any given $y \in Y^s$. If furthermore the KKT conditions (6) are satisfied with strict complementarity then the function $z(y)$ is continuously differentiable in the vicinity of the parameter $y^0$ [Jit84]. Due to Assumption 1 the upper-level objective function $F(y, z)$ is differentiable with respect to $z$. Therefore, $F(y, z(y))$ is differentiable with respect to $y \in$ ri $Y^s$. ■

**Proposition 5** *Assume that*

1. *Assumptions 1 and 2 hold,*

2. *the gradient $\nabla_z f(y, z)$ is linear in $y$ and $z$.*

*Then the segment $Y^s$ is convex and compact.*

**Proof.** With Assumption 2 the system (6) of Karush Kuhn Tucker conditions characterizes $z$ as the optimal lower-level response to the upper-level parameter $y \in Y^s$. According to Definition 2 this system defines the constraints of the segment $Y^s$. Hence, the set $Y^s$ is convex if the involved equality constraints are linear and the inequality constraints convex. These conditions are given with Assumption 1 and condition 2. The compactness of the segment $Y^s$ follows directly from the compactness of $Y$ and from Definition 2. ■

With the conditions of this proposition the Karush Kuhn Tucker conditions on the follower's problem (2) represent Linear Complementarity Constraints [CPS92] and problem (3) is a MPEC-convexly constrained program [Fle05]. The analysis of the stochastic programming problems in the subsequent section is restricted to this problem type. We refine therefore Assumption 1 as follows.

**Assumption 3** *The upper-level objective function $F(y, z)$ is convex in $y$ and $z$ and at least $C^2$, the gradient of the lower-level objective function $\nabla_z f(y, z)$ is linear in $y$ and $z$. The upper-level constraints $G_i(y, z), i = 1, ..., p$, are convex in $y$ and $z$ and at least $C^1$. The lower-level constraints $g_j(y, z), j = 1, ..., q$, are linear in $y$ and $z$.*

The Karush Kuhn Tucker conditions (6) may not be satisfied with strict complementarity at a point $y^b$ on the boundary of a segment $Y^s$ to an adjacent segment. Therefore differentiability of the function $F(y, z(y))$ at the boundary between adjacent segments can not be guaranteed.

As indicated above, problem (1) – (2) can be decomposed into a family of convex one-level problems by partitioning the inducible region $IR$ into segments $Y^s$.

$$\begin{aligned} &\min_{y,z,\lambda} F(y, z) \\ &E(y, z, \lambda) \leq 0 \\ &e(y, z, \lambda) = 0 \end{aligned} \tag{7}$$

with suitably defined constraints $E(y, z, \lambda)$ and $e(y, z, \lambda)$. These subproblems can then be solved separately.

In our subsequent discussions we need a stationarity condition. For the sake of transparency we state here the concept of strong stationarity for the deterministic problem and extend it to stochastic programming problems in Section 3. The deterministic formulation is based on results for general MPECs [Fle05, SS00, Ye05] and adapted to problem (3) taking into account the linearity of the lower-level constraints (Assumption 3). If not otherwise stated, the gradient is taken here with respect to $(y, z)$.

**Definition 6** *(Strong stationarity) A feasible point* $x^0 = (y^0, z^0, \lambda^0)$ *is called strongly stationary if there exists a vector of multipliers* $(\kappa, \nu, \zeta, \xi)$ *such that*

$$\nabla F(y^0, z^0) + \kappa^T \nabla G(y^0, z^0) + \nabla(\nabla_z f(y^0, z^0)^T \nu) + \zeta^T \nabla g(y^0, z^0) = 0 \tag{8a}$$

$$\nabla_z g(y^0, z^0)^T \nu - \xi^T = 0 \tag{8b}$$

$$\kappa^T G(y^0, z^0) = 0 \tag{8c}$$

$$G(y^0, z^0) \leq 0$$

$$\kappa \geq 0$$

$$\xi_i = 0, \qquad i \in \overline{I}_L \cap I_C$$

$$\zeta_i = 0, \qquad i \in I_L \cap \overline{I}_C$$

$$\zeta_i \geq 0, \xi_i \geq 0, \qquad i \in I_L \cap I_C$$

# 3 Stochastic two-stage problems with bilevel structure

In this section, we study two-stage stochastic programming problems with a bilevel structure. This structure is similar to problem (1) – (2). Recall that the leader's uncertainty can be divided into two types: the uncertainty about system parameters and that in his belief about other decision makers. The first type of uncertainty is expressed by a vector $\omega \in \Omega$ of random variables with a given probability distribution and then taken into account by a second-stage problem at the upper level. The second type is treated separately by the bilevel structure and the follower's response can be determined by solving the lower-level decision problem.

In the first variant of such a stochastic programming problem, only the leader can accommodate a recourse decision. Assuming the case of simple recourse, we state sufficient optimality conditions and develop a solution algorithm. Then the problem formulation is extended to two model variations where the follower's decision problem involves a second-stage decision. We show that also these models can be reformulated similarly to the first problem. Hence the presented solution algorithm can be applied also to the more complex problem versions.

Consider at first the following formulation with a two-stage stochastic programming problem in the upper level and a one-stage stochastic programming problem in the lower level:

$$\min_{y_1 \in Y_1} \{F_1(y_1, z_1^*) + \mathbb{E}_\omega Q(y_1, \omega)\} \tag{9a}$$

$$G(y_1, z_1^*) \leq 0 \tag{9b}$$

$$z_1^* = \arg \min_{z_1 \in Z_1} \mathbb{E}_\omega f_1(y_1, z_1, \omega) \tag{9c}$$

$$\mathbb{E}_\omega g(y_1, z_1, \omega) \leq 0 \tag{9d}$$

$$Q(y_1, \omega) = \min_{y_2 \in Y_2} F_2(y_1, y_2, \omega) \tag{9e}$$

$$W_1(\omega) y_2 = h_1(\omega) - T_1(\omega) y_1 \tag{9f}$$

with $Q : \mathbb{R}^n \times \Omega \to \mathbb{R}^1$, $F_2 : \mathbb{R}^n \times \mathbb{R}^{n_2} \times \Omega \to \mathbb{R}^1$, $W_1 \in \mathbb{R}^{n_2} \times \mathbb{R}^{p_2}$, $T_1 \in \mathbb{R}^n \times \mathbb{R}^{p_2}$, $h_1 \in \mathbb{R}^{p_2}$ and $Y_2 \subseteq \mathbb{R}^{n_2}$.

Problem (9) represents the simplest formulation of a two-stage stochastic programming problem with bilevel structure. In order to find a first-stage decision $y_1$, the leader takes into account his recourse decision $y_2$ and predicts the follower's response $z_1^*$. We assume that the influence of the follower's decisions is not strong enough to be regarded in the second stage. Therefore this response is not included into the recourse problem here.

The recourse problem (9e) – (9f) affects the first-stage problem only through the leader's objective function. If the follower's decision problem (9c) – (9d) satisfies Assumption 2, problem (9) is therefore equivalent to a one-level reformulation similar to (3).

$$\min_{y_1,z_1,\lambda} \{F_1(y_1,z_1) + \mathbb{E}_\omega Q(y_1,\omega)\} \tag{10a}$$

$$\mathbb{E}_\omega\{\nabla_{z_1} f_1(y_1,z_1,\omega) + \lambda^T \nabla_{z_1} g(y_1,z_1,\omega)\} = 0$$

$$\lambda^T \mathbb{E}_\omega g(y_1,z_1,\omega) = 0 \tag{10b}$$

$$G(y_1,z_1) \le 0$$

$$\mathbb{E}_\omega g(y_1,z_1,\omega) \le 0$$

$$Q(y_1,\omega) = \min_{y_2 \in Y_2} F_2(y_1,y_2,\omega) \tag{10c}$$

$$W_1(\omega)y_2 = h_1(\omega) - T_1(\omega)y_1 \tag{10d}$$

This problem is a stochastic programming problem with recourse and the complementarity constraint (10b). In order to apply a stationarity concept to this problem, we need the convexity and differentiability of the leader's objective function $F_1(y_1,z_1) + \mathbb{E}_\omega Q(y_1,\omega)$. The follower's objective function $\mathbb{E}_\omega f_1(y_1,z_1,\omega)$ is convex and differentiable due to Assumption 3.

**Proposition 7** *If the function $F_1(y_1,z_1)$ is convex in $y_1$ and $F_2(y_1,y_2,\omega)$ is convex in $y_1$ and $y_2$ for all $\omega$ then the function $F_1(y_1,z_1) + \mathbb{E}_\omega Q(y_1,\omega)$ is convex in $y_1$. It is differentiable with respect to $y_1$ almost everywhere. If the random variable $\omega$ is absolutely continuously distributed, then the function $F_1(y_1,z_1) + \mathbb{E}_\omega Q(y_1,\omega)$ is continuously differentiable with respect to $y_1$.*

**Proof.** See for example Birge and Louveaux [BL97]. ■

We collect the assumptions of this proposition as follows.

**Assumption 4** *The function $F_1(y_1,z_1)$ is convex in $y_1$, $F_2(y_1,y_2,\omega)$ is convex in $y_1$ and $y_2$ for all $\omega$ and the random variable $\omega$ is absolutely continuously distributed.*

These considerations facilitate an adaptation of the strong stationarity conditions (8) to problem (10). Under a constraint qualification, strong stationarity has been established as a necessary optimality condition by Scheel and Scholtes [SS00]. However, for MPEC-convexly constrained problems with a convex objective function strong stationarity is even a sufficient optimality condition [Fle05].

In the following, we denote the optimal recourse decision by $y_2^0(\omega)$ and the associated Lagrange multiplier by $v^{R0}(\omega)$ for given first-stage decision $y_1^0$ and observation $\omega$ of the random variable.

**Theorem 8** *Assume that*

1. *Assumptions 2 – 4 hold,*
2. *the point* $x^0 = (y_1^0, z_1^0, \lambda^0)$ *is feasible for problem* (9)*,*
3. *there exists a vector of multipliers* $(\kappa, \nu, \zeta, \xi)$ *such that*

$$\nabla F_1(y_1^0, z_1^0) + \mathbb{E}_\omega\{\nabla F_2(y_1^0, y_2^0(\omega), \omega) - v^{R0}(\omega)T_1(\omega)\} + \kappa^T \nabla G(y_1^0, z_1^0)$$
$$+\mathbb{E}_\omega\{\nabla(\nabla_{z_1} f_1(y_1^0, z_1^0)^T \nu) + \zeta^T \nabla g(y_1^0, z_1^0, \omega)\} = 0 \tag{11a}$$
$$\mathbb{E}_\omega \nabla_z g(y_1^0, z_1^0, \omega)^T \nu - \xi^T = 0 \tag{11b}$$
$$\kappa^T G(y_1^0, z_1^0) = 0 \tag{11c}$$
$$G(y_1^0, z_1^0) \leq 0$$
$$\kappa \geq 0$$
$$\xi_i = 0, \qquad i \in \overline{I}_L \cap I_C$$
$$\zeta_i = 0, \qquad i \in I_L \cap \overline{I}_C$$
$$\zeta_i \geq 0, \xi_i \geq 0, \qquad i \in I_L \cap I_C$$

*Then the point* $x^0$ *is a local optimal solution of problem* (9)*.*

**Proof.** Under Assumption 2 the stochastic programming problem with bilevel structure (9) is equivalent to the stochastic one-level problem (10). Therefore, if $x^0 = (y^0, z^0, \lambda^0)$ is feasible for (9), it is also feasible for (10).
Due to Assumption 4 the recourse function $\mathbb{E}_\omega Q(y_1^0, \omega)$ is differentiable at $y_1^0$. For given decision $y_1^0$, observation $\omega$ and recourse decision $y_2^0(\omega)$ the gradient of the recourse function with respect to $(y_1, z_1)$ can be determined using the Lagrangian function of the recourse problem

$$\begin{aligned}\nabla \mathbb{E}_\omega Q(y_1^0, \omega) &= \nabla \mathbb{E}_\omega L^R(y_1^0, y_2^0(\omega), v^{R0}(\omega)) \\ &= \mathbb{E}_\omega\{\nabla F_2(y_1^0, y_2^0(\omega), \omega) - v^{R0}(\omega)T_1(\omega)\}\end{aligned} \tag{12}$$

Now, keeping in mind that in problem (9) the leader's objective function is $F_1(y_1^0, z_1^0) + \mathbb{E}_\omega Q(y_1^0, \omega)$ and the follower's objective function is $\mathbb{E}_\omega f_1(y_1^0, z_1^0)$, system (11) represents an adaptation of the strong stationarity conditions (8) to problem (10).
If Assumption 3 holds, problem (10) has a MPEC-convexly constrained structure. Furthermore, with Assumption 4, it has a convex objective function. Then the point $x^0 = (y_1^0, z_1^0, \lambda^0)$ is a local optimum of problem (10) ([Fle05], Theorem 4.7) and thus also of the original problem (9). ■

Now, we direct our attention to a solution method for problem (9). We apply the partitioning strategy outlined in the previous section. The original problem (9) is partitioned into a family of stochastic one-level problems described by segments of the upper-level domain. Then, using a stochastic quasi-gradient method [Erm88, Gai88, Gai04], a stationary point on a segment is found. Finally, the optimality of this point with regard to the original problem (9) is tested and possibly the search is continued on a new segment.

This strategy implies that on a segment $Y^s$ the following problem is solved.

$$\min_{y_1,z_1,\lambda} \{F_1(y_1,z_1) + \mathbb{E}_\omega Q(y_1,\omega)\} \tag{13a}$$

$$\mathbb{E}_\omega E(y_1,z_1,\lambda,\omega) \le 0 \tag{13b}$$

$$\mathbb{E}_\omega e(y_1,z_1,\lambda,\omega) = 0 \tag{13c}$$

$$Q(y_1,\omega) = \min_{y_2\in Y_2} F_2(y_1,y_2,\omega) \tag{13d}$$

$$W_1(\omega)y_2 = h_1(\omega) - T_1(\omega)y_1 \tag{13e}$$

$$(y_1,z_1,\lambda) \in Y_1 \times Z_1 \times \mathbb{R}^q_+$$

with

$$E(y_1,z_1,\lambda,\omega) = \begin{pmatrix} g_i(y_1,z_1,\omega), & i \in \overline{I}^s_C \cap I^s_L \\ -\lambda_i, & i \in I^s_C \cap \overline{I}^s_L \\ G(y_1,z_1) & \end{pmatrix} \tag{14}$$

$$e(y_1,z_1,\lambda,\omega) = \begin{pmatrix} \nabla_z f_1(y_1,z_1,\omega) + \lambda^T \nabla_z g(y_1,z_1,\omega) & \\ g_i(y_1,z_1,\omega), & i \in I^s_C \\ \lambda_i, & i \in I^s_L \end{pmatrix} \tag{15}$$

**Algorithm 1: Find local optimum among stationary points on segments.**

Step 0. (Initialisation) Find an initial upper-level decision $y_1^0$, set $s = 0$.

Step 1. (Determination of segment) Solve the lower-level problem (9c) – (9d) with the parameter $y_1^s$. This gives the optimal lower-level response $z_1^s = z(y_1^s)$, the associated Lagrange multipliers $\lambda^s$, the index set $I^s_C$ of active lower-level constraints and the index set $I^s_L$ of zero Lagrange multipliers.

Step 2. (Iteration) utilizing the initial point $x^s = (y_1^s, z_1^s, \lambda^s)$, solve problem (13) – (15), for example by Algorithm 2. A stationary solution $\overline{x}^s = (\overline{y}_1^s, \overline{z}_1^s, \overline{\lambda}^s)$ is obtained.

Step 3. (Optimality test) If the point $\overline{x}^s = (\overline{y}_1^s, \overline{z}_1^s, \overline{\lambda}^s)$ with the recourse decision $\overline{y}_2^s$ satisfies the optimality conditions (11) go to Step 5.

Step 4. (Perturbation into feasible descent direction)
Choose a descent direction $d$ which is feasible on an adjacent segment. Perturb $\overline{y}_1^s$ into that direction

$$y_1^{s+1} = \overline{y}_1^s + \beta d$$

with small $\beta > 0$. Set $s = s + 1$ and go to Step 1.

Step 5. (Termination) The point $\overline{y}_1^s$ with the optimal lower-level response $\overline{z}_1^s$ and the recourse decision $\overline{y}_2^s$ is a local optimal solution of problem (9).

**Remark 9** *1. Determination of an initial point in Step 0. An initial point $y_1^0$ is assumed to be feasible together with the response $z_1^0$, i.e. it may be any $x^0 = (y_1^0, z_1^0, \lambda^0) \in Y_1 \times Z \times R_+^q$ satisfying*

$$\mathbb{E}_\omega\{\nabla_z f(y_1^0, z_1^0, \omega) + (\lambda^0)^T \nabla_z g(y_1^0, z_1^0, \omega)\} = 0 \tag{16a}$$
$$(\lambda^0)^T \mathbb{E}_\omega g(y_1^0, z_1^0, \omega) = 0 \tag{16b}$$
$$\mathbb{E}_\omega g(y_1^0, z_1^0, \omega) \leq 0 \tag{16c}$$
$$G(y_1^0, z_1^0) \leq 0 \tag{16d}$$

*Here, conditions* (16a) – (16c) *characterize $z_1^0$ as optimal lower-level response and* (16d) *denotes the upper-level feasibility.*
*With this initial point $x^0$ the response $z_1^0$ and the Lagrange multipliers $\lambda^0$ are already determined such that Step 1 in Algorithm 1 is basically completed. It remains only to determine the index sets $I_C^s$ and $I_L^s$. If these sets change in the close vicinity of an initial point $y_1^0$ this means that this point is located on the boundary of several adjacent segments. In such a case an initial segment may be chosen arbitrarily among these segments.*

2. *Determination of lower-level response $z_1^s$ and Lagrange multipliers $\lambda^s$ for given $y_1^s$. Under assumption 3, problem* (9c) – (9d) *with the parameter $y_1^s$ represents a common stochastic programming problem with linear constraints. It can be solved using any standard SP approach.*

3. *Determination of feasible descent directions in Step 4. A failure of the optimality test of the stationary point $\overline{y}_1^s$ implies that this point is on the boundary of the segment $Y^s$ and there may exist directions of descent into an adjacent segment, say $Y^t$. Then the sets of active lower-level constraints and of zero Lagrange multipliers associated to a response to the perturbed point $y_1^t = \overline{y}_1^s + \beta d$ change. The behavior of the upper-level objective function can not be evaluated without solving the lower-level problem. However, the perturbed point $y_1^t$ is in the relative interior of the segment $Y^t$ and it can be assumed that condition* (16b) *is satisfied with strict complementarity. Therefore a possible approach to find a feasible descent direction on another segment is the following.*

   *Test if for any index sets $I_C^t \neq I_C^s$ and $I_L^t \neq I_L^s$ the system*

$$\mathbb{E}_\omega\{\nabla_z f(\overline{y}_1^s + \beta d, z_1^t, \omega) + (\lambda^t)^T \nabla_z g(\overline{y}_1^s + \beta d, z_1^t, \omega)\} = 0 \tag{17a}$$
$$\mathbb{E}_\omega g_i(\overline{y}_1^s + \beta d, z_1^t, \omega) = 0, \quad i \in I_C^t \tag{17b}$$
$$\lambda_i^t > 0, \quad i \in I_C^t \cap \overline{I}_L^t \tag{17c}$$
$$\mathbb{E}_\omega g_i(\overline{y}_1^s + \beta d, z_1^t, \omega) < 0, \quad i \in \overline{I}_C^t \cap I_L^t \tag{17d}$$
$$\lambda_i^t = 0, \quad i \in I_L^t \tag{17e}$$
$$G(\overline{y}_1^s + \beta d, z_1^t) \leq 0 \tag{17f}$$
$$F(\overline{y}_1^s + \beta d, z_1^t) - F(\overline{y}_1^s, \overline{z}_1^s) + \mathbb{E}_\omega\{Q(\overline{y}_1^s + \beta d, \omega) - Q(\overline{y}_1^s, \omega)\} < 0 \tag{17g}$$

   *with small $\beta > 0$ has solutions $z^t, \lambda^t$ and $d \neq 0$. In this system, constraints* (17b) – (17e) *specify the strict complementarity, constraint* (17f) *ensures the upper-level* *feasibility and constraint* (17g) *the descent of the direction* $d$. *If such index sets* $I_C^t, I_L^t$ *exist, a new segment* $Y^{s+1} = Y^t$ *and a feasible direction of descent in that segment are found.*

4. *If it is complicated or impossible to calculate the expectations in systems* (16) *or* (17), *they can be approximated by various deterministic equivalent formulations obtained through a sufficiently large sample of observations of the random variable.*

Note that the number of segments grows exponentially with the number of constraints and decision variables of the lower-level problem (9c) – (9d). Therefore, Algorithm 1 has not been designed to conduct the search on all possible segments in order to find a global minimum. Rather, the search is, if necessary, extended only to segments adjacent to the currently studied segment until a local optimum is determined.

Problem (13) contains the stochastic equality constraints (15) which may complicate a solution by a projection method. Furthermore the second-stage problem (13d) – (13e) must be taken into account. Therefore a Lagrange multiplier method [NV77] is utilized solving the problem

$$\min_{x \in X} \max_{u \geq 0, v} \mathbb{E}_\omega L(x, u, v, \omega) \tag{18}$$

where

$$L(x, u, v, \omega) = F_1(x) + Q(x, \omega) + uE(x, \omega) + ve(x, \omega) \tag{19}$$

is the Lagrangian function of problem (13) with $x = (y_1, z_1, \lambda)$.

**Algorithm 2: Find stationary point in a segment utilizing Lagrangian.**

Step 0. (Initialisation) Set $k = 1$, the initial point $\hat{x}^0 = (\hat{y}_1^0, \hat{z}_1^0, \hat{\lambda}^0)$ is passed from Algorithm 1. The Lagrange multipliers $u^0 \in \mathbb{R}^{m+q}$ and $v^0 \in \mathbb{R}^{p+q}$ are associated to this point $\hat{x}^0$.

Step 1. (Recourse decision) Determine a sample $\{\omega^1, ..., \omega^{N_k}\}$ of observations of the random variable $\omega$.
For each observation $\omega^\nu$, $\nu = 1, ..., N_k$ solve the recourse problem (13d) – (13e) with the first-stage iterate $\hat{x}^k = (\hat{y}^k, \hat{z}^k, \hat{\lambda}^k)$ and obtain the recourse decision $y_2^{k,\nu} = y_2(\hat{x}^k, \omega^\nu)$, the Lagrange multipliers $v_R^{k,\nu} = v_R(\hat{x}^k, \omega^\nu)$ and the recourse function $Q(\hat{x}_1^k, \omega^\nu)$.

Step 2. (Objective function) Calculate an approximation $\tilde{F}(\hat{y}_1^k, \hat{z}_1^k)$ of the objective function and the estimation $\overline{F}^k$

$$\tilde{F}(\hat{y}_1^k, \hat{z}_1^k) = F_1(\hat{y}_1^k, \hat{z}_1^k) + \sum_{\nu=1}^{N_k} Q(\hat{x}_1^k, \omega^\nu)$$

$$\overline{F}^k = \frac{1}{k} \sum_{i=1}^{k} \tilde{F}(\hat{y}_1^i, \hat{z}_1^i)$$

Step 3. (Search direction and step size) Determine search directions

$$\xi_x^k = \nabla_x F_1(\hat{x}^k) + \frac{1}{N_k} \sum_{\nu=1}^{N_k} \Big( u^k \nabla_x E(\hat{x}^k) + v^k \nabla_x e(\hat{x}^k, \omega^\nu) + \nabla_x F_2(\hat{x}^k, y_2^{k,\nu}, \omega^\nu) - v_R^{k,\nu} T_1(\omega^\nu) \Big) \quad (20)$$

$$\xi_u^k = \frac{1}{N_k} \sum_{\nu=1}^{N_k} E(\hat{x}^k, \omega^\nu) \quad (21)$$

$$\xi_v^k = \frac{1}{N_k} \sum_{\nu=1}^{N_k} e(\hat{x}^k, \omega^\nu) \quad (22)$$

and step sizes $\alpha_x^k$, $\alpha_u^k$ and $\alpha_v^k$ satisfying the conditions

$$\begin{aligned} &\alpha_x \to 0^+, \quad \sum_{k=1}^{\infty} \alpha_x^k = \infty, \quad \sum_{k=1}^{\infty} (\alpha_x^k)^2 < \infty \\ &\alpha_u \to 0^+, \quad \sum_{k=1}^{\infty} \alpha_u^k = \infty, \quad \sum_{k=1}^{\infty} (\alpha_u^k)^2 < \infty \\ &\alpha_v \to 0^+, \quad \sum_{k=1}^{\infty} \alpha_v^k = \infty, \quad \sum_{k=1}^{\infty} (\alpha_v^k)^2 < \infty \\ &\frac{\alpha_x^k}{\alpha_u^k} \to 0, \quad \frac{\alpha_x^k}{\alpha_v^k} \to 0 \end{aligned} \quad (23)$$

Step 4. (Update) Determine new iterates for the upper-level decision $x$ and the Lagrange multipliers $u$ and $v$:

$$\begin{aligned} \hat{x}^{k+1} &= \Pi_X(\hat{x}^k - \alpha_x^k \xi_x^k) \\ \hat{u}^{k+1} &= \max\{0, \hat{u}^k + \alpha_u^k \xi_u^k\} \\ \hat{v}^{k+1} &= \hat{v}^k + \alpha_v^k \xi_v^k \end{aligned}$$

where the operator $\Pi_X$ denotes the projection on the feasible area $X = Y_1 \times Z_1 \times \boldsymbol{R}_+^q$.
Set $k = k + 1$.

Step 5. (Convergence) If a convergence test is satisfied, for example if

$$|\overline{F}^{k-j} - \overline{F}^{k-j-1}| \leq \varepsilon_c, \quad \forall\, j = 0, ..., n$$

for $k \geq n + 1$ with given precision $\varepsilon_c$ and test horizon $n \geq 0$, go to Step 6.
Otherwise go to Step 1.

Step 6. (Termination) The point $\overline{x}^s = \hat{x}^k$ is a stationary solution of problem (13), i.e. $\overline{x}^s$ is stationary on the segment $Y^s$.

**Remark 10** *1. Sample size $N_k$ in Step 1. The choice of a truly efficient sample size $N_k$ is complicated. A large sample size slows down the progress of the algorithm and is not required at iterates obviously not in the vicinity of a stationary point. On the other hand, the utilization of a small sample near a stationary point entails too much imprecision. This suggests to perform tests if the current iterate is in the vicinity of a stationary point and to adapt the number $N_k$ accordingly. A more detailed discussion of this aspect is provided for example in [Gai88].*

2. *Convergence test in Step 5. Since the random parameters are approximated by a sample of observations there may occur periods with apparently stationary iterates which obviously are not optimal. Especially if such a period occurs during the first iteration steps the estimation $\overline{F}^k$ seems to converge. In order to avoid the termination of the algorithm in such a case the convergence test evaluates the estimation over a horizon of $n$ iteration steps. (See also the implementation example in Section 4.)*

**Theorem 11** *(Convergence of Algorithm 2) Assume that*

1. *Assumptions 2 – 4 hold,*
2. *the search directions $\xi_x^k$, $\xi_u^k$ and $\xi_v^k$ are defined by* (20) – (22),
3. *the step sizes $\alpha_x^k$, $\alpha_u^k$ and $\alpha_v^k$ satisfy the conditions* (23).

*Then Algorithm 2 converges with probability 1 to the vicinity of a stationary point $\overline{x}^s$ of problem* (13).

**Proof.** Under Assumptions 3 and 4 the objective function of problem (13) is convex and continuously differentiable in $y_1$ and $z_1$. Due to Assumption 2 the convex optimization problem (13) is equivalent to the Lagrangian saddle point formulation (18). This problem is solved by Algorithm 2 utilizing a Lagrange multiplier method.
In order to determine the search directions an estimate of the subgradient of the recourse function is needed. For an iterate $\hat{x}^k$ and an observation $\omega^\nu$ of the random variable such an estimate is for example the gradient with respect to $x$ of the Lagrangian of the recourse problem:

$$\nabla_x L^R(\hat{x}^k, y_2^{k,\nu}, v_R^{k,\nu}, \omega^\nu) = \nabla_x F_2(\hat{x}^k, y_2^{k,\nu}, \omega^\nu) - v_R^{k,\nu} T_1(\omega^\nu)$$

Taking now into account that the inequality constraints $E(x)$ are deterministic, the search directions (20) – (22) satisfy the stochastic quasi-gradient conditions

$$\begin{aligned} I\!\!E_\omega\{\xi_x^k|\hat{x}^0, ..., \hat{x}^k\} &= \nabla_x I\!\!E_\omega L(\hat{x}^k, \hat{u}^k, \hat{v}^k, \omega) \\ I\!\!E_\omega\{\xi_v^k|\hat{x}^0, ..., \hat{x}^k\} &= \nabla_v I\!\!E_\omega L(\hat{x}^k, \hat{u}^k, \hat{v}^k, \omega) \\ I\!\!E_\omega\{\xi_u^k|\hat{x}^0, ..., \hat{x}^k\} &= \nabla_u I\!\!E_\omega L(\hat{x}^k, \hat{u}^k, \hat{v}^k, \omega) \end{aligned}$$

With the step size conditions (23) Algorithm 2 converges then with probability 1 to the vicinity of a stationary point $\overline{x}^s$ of problem (13) [Erm88]. ■

**Theorem 12** *Suppose that*

1. *Assumptions 2 – 4 hold,*
2. *the search directions* $\xi_x^k$, $\xi_u^k$ *and* $\xi_v^k$ *are defined by* (20) – (22)*,*
3. *the step sizes* $\alpha_x^k, \alpha_u^k$ *and* $\alpha_v^k$ *satisfy the conditions* (23)*,*
4. *the optimality test in Algorithm 1 utilizes the optimality conditions* (11).

*Then Algorithm 1 utilizing Algorithm 2 stops with probability 1 at a point in the vicinity of a local minimum of problem* (9).

**Proof.** Under conditions 1. and 2. Algorithm 2 stops with probability 1 at a point $\overline{x}^s$ in the vicinity of a stationary point of problem (13) on the segment $Y^s$. If this point satisfies the optimality conditions (11), a local minimum of problem (9) is found and Algorithm 1 stops.
If the optimality conditions are not satisfied, there exist feasible descent directions at the point $\overline{x}^s$. Since the point $\overline{x}^s$ is in the vicinity of a stationary point on the segment $Y^s$, there exist only directions of descent into adjacent segments. These segments are characterized by index sets $I_C^t \neq I_C^s$ and $I_L^t \neq I_L^s$. A direction $d$ is therefore found as solution of system (17). The point $\overline{x}^s$ is perturbed into this direction $d$ and the search is repeated on the new segment $Y^{s+1}$ with the initial point $(y^{s+1}, z^{s+1}, \lambda^{s+1})$. The number of the segments and thus of convex subproblems is finite. Therefore, Algorithm 1 stops with probability 1 after a finite number of steps at a point in the vicinity of a local optimum of problem (9). ■

Algorithm 2 stops at a point in the vicinity of a stationary point. Furthermore, the exact calculation of the expectation in equation (11a) of the optimality conditions may be difficult or impossible such that a sample of observations should be used (cf. Remark 9, 4.) Therefore the optimality conditions should possibly be verified not exactly but rather within certain tolerance bounds. Then, however, Algorithm 1 stops only with probability in the vicinity of a local optimal solution of problem (9).

Now the two-stage problem (9) is extended by taking into account a reaction of the follower on changed conditions at the second stage. This means that a bilevel relationship between the actors' problems exists at each stage. The follower's second-stage decision $z_2^*$ represents a reaction on the changed conditions in a similar sense as the leader's recourse decision, i.e. it is a correcting action. However, contrary to the leader's problem, the follower cannot take this second-stage decision into account when making her first-stage decision. Therefore the follower's second-stage problem is not interpreted as recourse problem. Furthermore, we suppose that the leader's first-stage decision can directly influence the follower's second-stage decision. This reflects the case when some of the leader's first-stage decisions still are valid for the control of the follower's decisions, such as certain regulatory obligations on the follower. Such a model can be formulated as follows.

$$\min_{y_1 \in Y_1} \{F_1(y_1, z_1^*) + E_\omega Q(y_1, z_1^*, \omega)\} \tag{24a}$$

$$G(y_1, z_1^*) \leq 0 \tag{24b}$$

$$Q(y_1, z_1^*, \omega) = \min_{y_2 \in Y_2} F_2(y_1, y_2, z_2^*, \omega) \tag{24c}$$

$$W_1(\omega)y_2 = h_1(\omega) - T_1(\omega)y_1 - U_1(\omega)z_1^* - V_1(\omega)z_2^* \tag{24d}$$

$$z_1^* = \arg\min_{z_1 \in Z_1} E_\omega f_1(y_1, z_1, \omega) \tag{24e}$$

$$g(y_1, z_1) \leq 0 \tag{24f}$$

$$z_2^* = \arg\min_{z_2 \in Z_2} f_2(y_1, y_2, z_1^*, z_2, \omega) \tag{24g}$$

$$V_2(\omega)z_2 = h_2(\omega) - T_2(\omega)y_1 - U_2(\omega)z_1^* - W_2(\omega)y_2 \tag{24h}$$

where $Z_2 \subseteq R^{m_2}, Q : R^n \times R^{m_2} \times R^n \to R^1$, $f_2 : R^{2m_2} \times R^{2n_2} \to R^1$, $T_2 \in R^n \times R^{q_2}, U_1 \in R^m \times R^{p_2}, U_2 \in R^m \times R^{p_2}$, $V_1 \in R^{m_2} \times R^{p_2}, V_2 \in R^{m_2} \times R^{q_2}, W_2 \in R^{n_2} \times R^{q_2}$ and $h_2 \in R^{q_2}$. We assume that the follower's problems in both stages satisfy Assumption 2 and that the matrix $W_2(\omega)$ has full rank for any $\omega$.

The leader finds an optimal solution of his first-stage problem (24a) – (24b) taking into account the recourse problem (24c) – (24d). For this purpose he predicts the response $z_1^*$ to his first-stage decision $y_1$ and the response $z_2^*$ to his first-stage decision $y_1$ and to his recourse decision $y_2$. These responses can be determined by solving the follower's problems (24e) – (24f) and (24g) – (24h), respectively. The presence of inequality constraints in the follower's second-stage problem (24g) – (24h) would significantly complicate the analysis. This issue is addressed closer below.

**Theorem 13** *Suppose that*

1. *Assumptions 2 – 4 hold,*
2. *the second-stage objective function $f_2(y_1, y_2, z_1^*, z_2, \omega)$ of the follower is continuously differentiable in $z_2$,*
3. *the gradient $\nabla_{z_2} f_2(y_1, y_2, z_1^*, z_2, \omega)$ is linear in $y_1, y_2, z_1$ and $z_2$.*

*Then problem* (24) *can be formulated as a stochastic programming problem with a structure similar to problem* (9).

**Proof.** The follower's second-stage decision $z_2$ represents an optimal response to the leader's decisions at both stages. Due to Assumption 2 it can be substituted into the leader's recourse problem (24c) – (24d) utilizing the Karush Kuhn Tucker conditions on the second-stage lower-level problem (24g) – (24h). This results in a one-level formulation of the recourse problem:

$$Q(y_1, z_1, \omega) = \min_{y_2, z_2, \mu} F_2(y_1, y_2, z_2, \omega) \tag{25a}$$

$$\nabla_{z_2} f_2(y_1, y_2, z_1, z_2, \omega) + \mu^T V_2(\omega) = 0 \tag{25b}$$

$$W_1(\omega)y_2 + V_1(\omega)z_2 = h_1(\omega) - T_1(\omega)y_1 - U_1(\omega)z_1 \tag{25c}$$

$$W_2(\omega)y_2 + V_2(\omega)z_2 = h_2(\omega) - T_2(\omega)y_1 - U_2(\omega)z_1 \tag{25d}$$

Under condition 3. constraint (25b) can be expressed as

$$A(\omega)y_2 + B(\omega)z_2 + \mu^T V_2(\omega) = c^T(\omega) - D_1(\omega)y_1 - D_2(\omega)z_1 \tag{26}$$

with $A \in \boldsymbol{R}^{n_2} \times \boldsymbol{R}^{m_2}$, $B \in \boldsymbol{R}^{m_2} \times \boldsymbol{R}^{m_2}$, $c \in \boldsymbol{R}^{m_2}$, $D_1 \in \boldsymbol{R}^{n} \times \boldsymbol{R}^{m_2}$, $D_2 \in \boldsymbol{R}^{m} \times \boldsymbol{R}^{m_2}$. Resulting, constraints (25b) – (25d) can be collected in one linear constraint

$$\overline{W}(\omega)v_2 = \overline{h}(\omega) - \overline{T}(\omega)v_1$$

with

$$\overline{T}(\omega) = \begin{bmatrix} D_1(\omega) & D_2(\omega) \\ T_1(\omega) & U_1(\omega) \\ T_2(\omega) & U_2(\omega) \end{bmatrix} \tag{27}$$

$$\overline{W}(\omega) = \begin{bmatrix} A(\omega) & B(\omega) & V_2^T(\omega) \\ W_1(\omega) & V_1(\omega) & 0 \\ W_2(\omega) & V_2(\omega) & 0 \end{bmatrix} \tag{28}$$

$$\overline{h}(\omega) = (c(\omega), h_1(\omega), h_2(\omega))^T \tag{29}$$

$$v_1 = (y_1, z_1)^T \tag{30}$$

$$v_2 = (y_2, z_2, \mu)^T \in X_2 = Y_2 \times Z_2 \times \boldsymbol{R}_+^{q_2} \tag{31}$$

This way all information of the follower's second-stage problem is included into the recourse problem. Consequently, problem (24) is similar to problem (9). ■

Reformulating problem (24) in the vicinity of a feasible point $x^0 = (y_1^0, z_1^0, \lambda^0)$, a one-level two-stage stochastic programming problem is obtained which is similar to the formulation (10). Then the optimality conditions (11) can be applied to this problem in an equivalent way.

Finally, the following theorem states that Algorithm 1 coupled with Algorithm 2 can be utilized without modification for the solution of problem (24).

**Theorem 14** *Assume that*

1. *Assumptions 2 – 4 hold,*
2. *the search directions $\xi_x^k$, $\xi_u^k$ and $\xi_v^k$ are defined by* (20) – (22)*,*
3. *the step sizes $\alpha_x^k$, $\alpha_u^k$ and $\alpha_v^k$ satisfy the conditions* (23)*,*
4. *the optimality test in Algorithm 1 utilizes the optimality conditions* (11)*.*

*Then Algorithm 1 together with Algorithm 2 yields a local optimal solution of problem* (24).

**Proof.** Theorem 13 stated that problem (24) can be reformulated in such a way that it assumes the structure of problem (9):

$$\min_{y_1 \in Y_1} F_1(y_1, z_1^*) + \boldsymbol{E}_\omega Q(y_1, z_1^*, \omega) \tag{32a}$$

$$G(y_1, z_1^*) \le 0$$

$$z_1^* = \arg\min_{z_1 \in Z_1} \boldsymbol{E}_\omega f_1(y_1, z_1, \omega) \tag{32b}$$

$$g(y_1, z_1) \le 0$$

$$Q(y_1, z_1^*, \omega) = \min_{v_2 \in X_2} F_2(y_1, v_2, \omega) \tag{32c}$$

$$\overline{W}(\omega)v_2 = \overline{h}(\omega) - \overline{T}(\omega)(y_1, z_1) \tag{32d}$$

with $\overline{T}(\omega), \overline{W}(\omega), \overline{h}(\omega), v_2$ according to (27) – (29) and (31).
The second-stage decision variable $v_2$ of problem (32) comprises the leader's recourse decision $y_2$, the follower's second-stage decision $z_2$ as response to $y_2$ and the Lagrange multipliers $\mu$ associated to this response. Since the follower's second-stage response is unique for all recourse decisions of the leader, the variable $v_2$ can be determined as optimal solution of the recourse problem (32c) – (32d). Therefore, following the reasoning of Theorem 11, Algorithm 1 together with Algorithm 2 can be applied to problem (24) without modification. It yields a local optimal solution of the reformulation (32) and thus also a local optimal solution of the original problem (24). ■

The analysis of a stochastic programming problem with recourse and bilevel structure in the formulation (24) is considerably more complex when the second-stage lower-level constraints (24h) comprise inequality constraints. In this case the one-level reformulation (25) of the second-stage problem contains a complementarity constraint. Consequently, the leader's recourse problem does not satisfy the Slater constraint qualification for any given first-stage decision and random parameter. This, however, is a precondition for the convexity of the second-stage optimal value function $Q$ ([BL97], Theorem 32).
One may apply the described partitioning strategy also to this second-stage problem and thus combine first-stage and second-stage segments. However, then only locally optimal recourse decisions can be found and the problem of verifying the convexity of the function $Q$ still persists. We leave this issue for further research.

# 4 Numerical studies

This section provides the results of some numerical examples which shall demonstrate the viability of the presented approach and provide some indications about characteristics of the algorithm.

## 4.1 Model formulation

We apply the proposed solution method to an example from telecommunications which is a simplified version of the relationship described in [AGW06].

Both decision makers maximize their profits from the provision of a telecom service to a common customer population. The follower is lacking essential infrastructure (e.g. network capacity) and relies on the leader for access to such equipment. The customer demand depends on the decisions of both providers such that connecting upper-level constraints are present. At stage one, decisions about service prices, capacity price and the amount of leased capacity are made on the base of previous, deterministic data. At the second stage, the leader has the option to adapt to the observed environmental state by adding capacity for himself. Then, the following stochastic programming problem with bilevel structure and a recourse problem in the upper level can be formulated. It represents a version of model (24) with simple recourse, $W_1(\omega) = 1$, and no second-stage decisions of the follower, $V_1(\omega) = 0$. The leader's first and second-stage decision problems are

$$\max_{y_1} \; y_1^T C_{11} y_1 + y_1^T C_{12} z + d_{11}^T y_1 + d_{12}^T z + \mathbb{E}_\omega Q(y_1, z, \omega) \tag{33a}$$
$$A_1 y_1 + B_1 z + f_1 \leq 0 \tag{33b}$$
$$y_1 \in Y_1 \subset \boldsymbol{R}^3$$
$$Q(y_1, z, \omega) = \max_{y_2 \geq 0} \; q(\omega) y_2 \tag{33c}$$
$$y_2 = h_1(\omega) - T_1(\omega) y_1 - U_1(\omega) z \tag{33d}$$

The follower's response on a first-stage decision $y_1$ is found as an optimal solution of the following problem:

$$\max_{z} \; y_1^T C_{21} z + z^T C_{22} z + d_{21}^T y_1 + d_{22}^T z \tag{34a}$$
$$A_2 y_1 + B_2 z + f_2 \leq 0 \tag{34b}$$
$$z \in Z \subset \boldsymbol{R}^3$$

For a given initial point a segment $Y^s$ is described by the sets $I_L^s, I_C^s$ of zero Lagrange multipliers and active lower-level constraints. The second-stage problem is not affected by the partitioning since it has no bilevel structure. This results in the one-level reformulation

$$\min_{y_1, z, \lambda} F_1(y_1, z) + \mathbb{E}_\omega Q(y_1, z, \omega) \tag{35a}$$
$$E(y_1, z, \lambda) \leq 0 \tag{35b}$$
$$e(y_1, z, \lambda) = 0 \tag{35c}$$
$$Q(y_1, z, \omega) = \min_{y_2 \geq 0} q(\omega) y_2 \tag{35d}$$
$$y_2 = h_1(\omega) - T_1(\omega) y_1 - U_1(\omega) z \tag{35e}$$

with

$$E(y_1, z, \lambda) = \begin{pmatrix} A_1 y + B_1 z + f_1 & \\ A_{2i} y + B_{2i} z + f_{2i}, & i \in I_L^s \cap \overline{I}_C^s \\ -\lambda_i & i \in \overline{I}_L^s \cap I_C^s \end{pmatrix}$$

$$e(y_1, z, \lambda) = \begin{pmatrix} A_3 y_1 + B_3 z + C_3 \lambda + f_3 & \\ A_{2i} y + B_{2i} z + f_{2i}, & i \in I_C^s \\ \lambda_i & i \in I_L^s \end{pmatrix}$$

The model parameters are given as follows

$$B_1 = \begin{pmatrix} -r_{12} & 0 & 0 \\ dr_{12} & 1 & 0 \end{pmatrix} \qquad A_1 = \begin{pmatrix} r_{11} & 0 & 0 \\ -dr_{11} & 0 & -d \end{pmatrix} \qquad f_1 = \begin{pmatrix} -k_1 \\ dk_1 - b \end{pmatrix}$$

$$C_{11} = \begin{pmatrix} r_{11} & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad C_{12} = \begin{pmatrix} -r_{12} & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad d_{12} = \begin{pmatrix} -r_{12}q_1 \\ 0 \\ 0 \end{pmatrix}$$

$$B_2 = \begin{pmatrix} r_{22} & 0 & 0 \\ -dr_{22} & -1 & -d \end{pmatrix} \qquad A_2 = \begin{pmatrix} -r_{21} & 0 \\ dr_{21} & 0 \end{pmatrix} \qquad f_2 = \begin{pmatrix} -k_2 \\ dk_2 \end{pmatrix}$$

$$C_{21} = \begin{pmatrix} -r_{21} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad C_{22} = \begin{pmatrix} r_{22} & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad d_{21} = \begin{pmatrix} -q_2 r_{21} \\ 0 \\ 0 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} -r_{21} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad B_3 = \begin{pmatrix} 2r_{22} & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad C_3 = \begin{pmatrix} -r_{22} & dr_{22} \\ 0 & 1 \\ 0 & d \end{pmatrix}$$

$$d_{11} = (q_1 r_{11} - k_1, 0, q_1 + r_{13})^T \qquad d_{22} = (q_2 r_{22} - k_2, 0, q_2 + r_{23})^T$$
$$f_3 = (q_2 r_{22} - k_2, 0, q_2 + r_{23})^T$$

with

$$\begin{aligned} &r_{11} = r_{22} = 500, \;\; r_{12} = r_{21} = 250, \;\; r_{13} = r_{23} = 25 \\ &q_1 = q_2 = 1, \;\; k_1 = k_2 = 1300, \;\; d = 1, \;\; b = 2000 \\ &Y_1 = [0, 20] \times [0, 20] \times \boldsymbol{R}^1_+, \;\; Z = [0, 20] \times [0, 1000] \times \boldsymbol{R}^1_+ \end{aligned}$$

The random parameter is $\omega = [\tilde{r}_{11}, \tilde{r}_{12}, \tilde{r}_{14}, \tilde{k}_1]$ such that

$$q(\omega) = \tilde{r}_{14}, \; h(\omega) = \tilde{k}_1 - b, \; T(\omega) = (\tilde{r}_{11}, 0, 1)\,, \; U(\omega) = (-\tilde{r}_{12}, -1, 0)$$

We test the algorithm for two sets of random data which are uniformly distributed as follows

| | $\tilde{r}_{11}$ | $\tilde{r}_{12}$ | $\tilde{r}_{14}$ | $\tilde{k}_1$ |
|---|---|---|---|---|
| Slightly stochastic | $[490, 510]$ | $[240, 260]$ | $[5.8, 6.2]$ | $[1290, 1310]$ |
| Heavily stochastic | $[300, 700]$ | $[50, 450]$ | $[2, 10]$ | $[300, 2300]$ |

## 4.2 Implementation and results

For these first illustrative examples, the algorithm has been implemented in MATLAB utilizing the optimization toolbox. However, a more sophisticated and efficient implementation is under way which will allow for a speed-up and fine-tuning of the algorithm and, most importantly, for a comparison with other approaches and SMPEC test problems.

In order to decrease the computation time we employed two types of iteration steps. In a normal step only one observation of the random data is utilized for the calculations.

At regular intervals, a control step is performed with a sufficiently large sample of observations. Furthermore at such a step the step sizes are adjusted, either automatically or interactively. In the first case the step sizes are calculated according to a rule satisfying the conditions (23). An interactive step size adjustment allows the user to revise the step size according to his observations of the progress of the iteration. For more details on the step size strategy and for reasons for adopting such an interactive approach we refer to the discussion in [Gai88]. Note that with this strategy the step sizes can reach arbitrarily small values but do they not approach zero. Therefore only the vicinity of the optimal solution is reached and the interactive step size selection may be utilized as an indicator for a good automatic step size strategy.

The step sizes were determined according to the rules

$$\alpha_x^k = \frac{0.1}{1+s}, \qquad \alpha_u^k = \frac{0.1}{1+s^\gamma} \tag{36}$$

Alternatively, we tested the step size rule

$$\alpha_x^k = \frac{0.25}{2^s}, \qquad \alpha_u^k = \frac{0.25}{1.9^s}, \quad k \in [2^{s-1}, 2^s] \tag{37}$$

Here $k$ denotes the number of the current iteration step whereas $s$ is the iteration step at which the previous control step was performed. We used a regular review interval of 10 steps such that $s = \lceil k/10 \rceil$, the greatest integer which is smaller than or equal to $k/10$. The convergence test performed in the iteration evaluates the behavior of the estimation $\overline{F}^k$ over the previous three iteration steps. A stationary point $\overline{x}^s$ was identified as optimal when it was within a vicinity of 0.02 % of the actual optimum.

Generally, the iterates show a behavior typical for SQG methods: after a period with heavy oscillations the vicinity of the optimal solution is reached quite fast. From that point on the approximation improves only slowly, small oscillations in the vicinity of the optimum persist. More specific, four different sections can be distinguished for the tested problem. At first the iterates oscillate heavily between two clusters relatively far away from the optimum, possibly some periods with stable objective function values exist. In the second section, the oscillations shift slowly toward a further cluster in the vicinity of the optimum. A short section of consolidation follows. The variance of the oscillations decreases rapidly and the iterates begin to cluster in the vicinity of the optimal solution. Finally, the iterates oscillate in the vicinity of the optimum. One reason for these oscillations is the step size strategy. Especially due to the behavior of the iterates in the first two periods the estimation $\hat{F}$ of the objective function converges only very slowly. Typically the algorithm terminates because a predefined number of iteration steps was reached. The optimality conditions (11) are often not satisfied and the existence of feasible ascent directions is analyzed. This indicates that the convergence and optimality tests are too strict taken into account the stochasticity of the data or the sample size chosen for the test was too small. As a consequence, so-called approximate stationary points could not be identified. Here, a reasonable relaxation of the tests seems appropriate in order to recognize points in a close vicinity of an optimum.

Two sets of experiments were performed. The first set analyzed the segment $I_C^s = \{2, 5, 7\}$ which is found, for example, with the initial points $y_1^0 = (0, 3.6, 0)$ or $y_1^0 = (5.2, 6.2, 0)$. On this segment, the deterministic problem has a local optimum at $y_{1D}^* =$

$(2.53, 4.87, 0)$ with the recourse decision $y_{2D}^* = 0$, the follower's response $z_D^* = (3.87, 0, 0)$ and the objective function value $F_D^* = 2233.31$. Here, the step size strategy (36) proved quite efficient. In the case of slightly stochastic data the periods with heavy oscillations were small and a vicinity of 2 % of the optimal solution was reached after approximately 100 iteration steps and a precision of 0.5 % after further 20 steps. However, even after additional 150 steps the approximation did not increase significantly, the iterates where in a vicinity of about 0.3 % of the optimum. A similar behavior of the iterates can be observed in the case of more random data. Table 1 compares iterates obtained on this segment using highly stochastic and deterministic data.

Table 1: Heavily stochastic and deterministic data, $y_1^0 = (0, 3.6, 0)$

| | heavily stochastic data | | | deterministic data | | |
|---|---|---|---|---|---|---|
| $k$ | $y_1$ | $\hat{F}$ | $F^s$ | $y_1$ | $\hat{F}$ | $F^s$ |
| 1 | (0.00, 3.60, 0) | 650.00 | 650.00 | (0.00, 3.60, 0) | 650.00 | 650.00 |
| 2 | (5.20, 6.20, 0) | -325.00 | -1300.00 | (5.20, 6.20, 0) | -325.00 | -1300.00 |
| 3 | (5.20, 6.20, 0) | -650.00 | -1300.00 | (5.20, 6.20, 0) | -650.00 | -1300.00 |
| 4 | (-0.00, 3.60, 0) | -325.00 | 650.00 | (-0.00, 3.60, 0) | -325.00 | 650.00 |
| 5 | (-0.00, 3.60, 0) | -260.63 | -3.16 | (-0.00, 3.60, 0) | -130.00 | 650.00 |
| 6 | (-0.00, 3.60, 0) | -108.86 | 650.00 | (-0.00, 3.60, 0) | 0.00 | 650.00 |
| 7 | (5.20, 6.20, 0) | -279.02 | -1300.00 | (5.20, 6.20, 0) | -185.71 | -1300.00 |
| 8 | (5.20, 6.20, 0) | -406.65 | -1300.00 | (5.20, 6.20, 0) | -325.00 | -1300.00 |
| 9 | (5.20, 6.20, 0) | -505.91 | -1300.00 | (5.20, 6.20, 0) | -433.33 | -1300.00 |
| 10 | (-0.00, 3.60, 0) | -409.86 | 454.59 | (-0.00, 3.60, 0) | -325.00 | 650.00 |
| 100 | (3.65, 5.42, 0) | 560.52 | 1408.30 | (2.72, 4.96, 0) | 898.32 | 2159.00 |
| 250 | (2.99, 5.09, 0) | 1439.17 | 2008.21 | (2.53, 4.87, 0) | 1699.36 | 2233.31 |
| 500 | (2.83, 5.01, 0) | 1815.00 | 2104.29 | (2.53, 4.87, 0) | 1966.33 | 2233.31 |
| 1000 | (2.58, 4.89, 0) | 2007.75 | 2216.37 | (2.53, 4.87, 0) | 2099.82 | 2233.31 |
| 1500 | (2.87, 5.03, 0) | 2077.38 | 2081.52 | (2.53, 4.87, 0) | 2144.32 | 2233.31 |
| 2000 | (2.36, 4.78, 0) | 2114.53 | 2278.31 | (2.53, 4.87, 0) | 2166.56 | 2233.31 |

However, the second set of experiments shows that the good performance of rule (36) can not be generalized. Choosing the initial point $y_1^0 = (0.83, 0, 210)$, the iteration is conducted on the segment $I_C^s = \{2, 7\}$. The local optimum of the deterministic problem is $y_{1D}^* = (2.53, 0.87, 0)$ with the recourse decision $y_{2D}^* = 0$, the follower's response $z_D^* = (1.87, 1000, 0)$ and the objective value $F_D^* = 1333.29$. Here, strategy (36) shows a weaker performance. After a few large initial oscillations a long period of about 300 steps follows with quite stable iterates. During the next 100 steps the vicinity of the optimum is approached with only a few oscillations. Finally, the iterates oscillate in the vicinity of the optimum. For the case of low stochasticity, Figure 1 depicts a typical behavior of the iterates for this strategy on both segments.

With the second step size rule (37), the algorithm performs slightly better for the initial point $y_1^0 = (0.83, 0, 210)$, but the performance is worse for the initial point $y_1^0 = (0, 3.6, 0)$ (see Figure 2). This observation underlines that a step size strategy which performs equally well for all problems can hardly be found. Rather, at first the algorithm may be run tentatively in interactive mode in order to obtain a conjecture for a good automatic strategy. Such an automatic strategy can for example be chosen from a toolbox containing several alternatives.

A more sophisticated and efficient implementation will allow for a speed-up and fine-

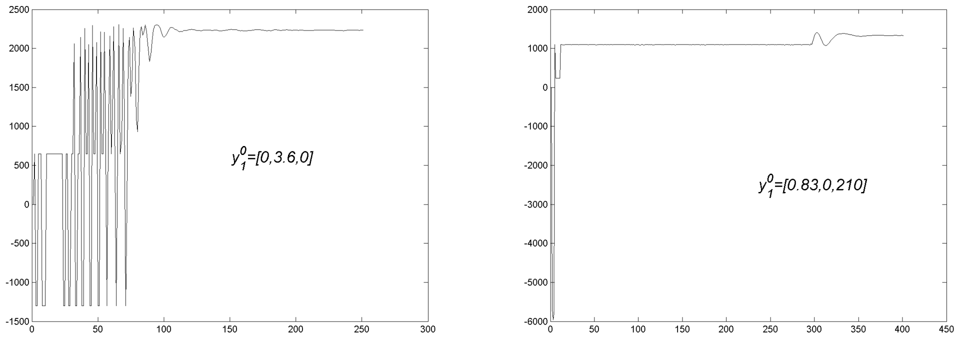


Figure 1: Estimated values of objective function, step size rule (36)

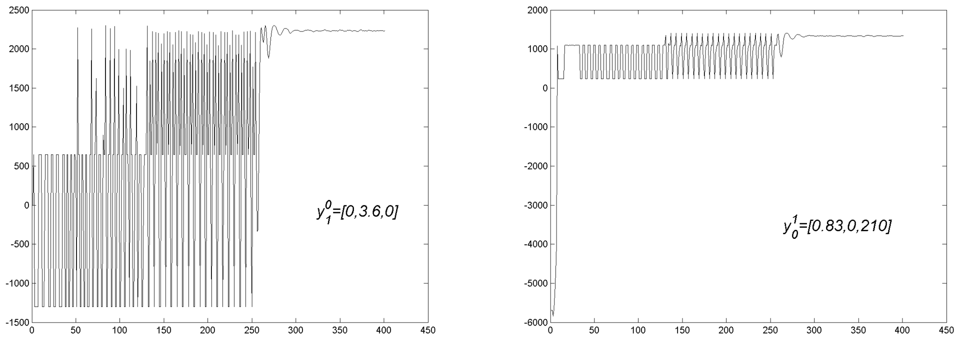


Figure 2: Estimated values of objective function; step size rule (37)

tuning of the proposed algorithm. For example, the test on existence of descent directions is now performed on all index sets. A first improvement may exclude the index sets of already visited segments from this search. Furthermore, a mechanism for identifying neighbor segments through their index sets will restrict the search to only a few segments at each iteration. The tests on stationarity and optimality may be improved using, for example, results by Bayraksan and Morton [BM06] on the quality of solutions of stochastic programming problems. Finally, a parallelization of the method may contribute to a considerable reduction of computation time.

# 5 Conclusions

We studied several formulations of stochastic programming problems with recourse and bilevel structure where connecting upper-level constraints are present. A strong stationarity concept has been stated which, under some assumptions on the considered stochastic programming problems, establishes sufficient conditions for optimality. An algorithm for the solution of the presented problem type has been developed utilizing a two-step solution process. This is due to the reformulation of the problems to MPEC-type one-level problems and, hence, the possible nonconvexity of the inducible region caused by the connecting upper-level constraints. We proved that, under certain conditions on the involved functions, a point in the vicinity of a local optimal solution of the originally studied problems is attained with probability 1. Tentative numerical experiments testify to a reasonable numerical efficiency of the proposed approach.

Future research may include more complex multiperiod problems. For example, the follower's second-stage problem may represent a recourse problem instead of the two-stage relationship studied now or it may contain inequality constraints. Another conceivable extension takes into account that the leader's perception of the follower's decision process may be imperfect. This means that the leader will obtain certainty about the actually implemented response only at the end of the first stage. Such a consideration of the uncertainty about the lower-level decision process is especially important for the analysis of agency problems. Finally, also extensions to problems involving multiple followers appear natural. Then, also the character of the interactions between these decision makers, such as Nash game or further Stackelberg game relations, and their effect on the model properties must be taken into account. Another field for further research may be concerned with enhancing and fine-tuning the proposed solution approach, following the lines of the discussion in Section 4.

# References

[AGW06] J.A. Audestad, A.A. Gaivoronski, and A.S. Werner. Extending the stochastic programming framework for the modeling of several decision makers: pricing and competition in the telecommunication sector. *Annals of Operations Research*, 142(1):19–39, 2006.

[BGL04a] S.I. Birbil, G. Gürkan, and O. Listes. Simulation-based solution of stochastic mathematical programs with complementarity constraints: sample-path analysis. Research report ers-2004-016-lis, Erasmus University, 2004.

[BGL04b] S.I. Birbil, G. Gürkan, and O. Listes. Solving stochastic mathematical programs with complementarity constraints using simulation. In R.I. Ingalls, M.D. Rossetti, J.S Smith, and B.A. Peters, editors, *Proceedings of the 2004 Winter Simulation Conference*, pages 550–557, 2004.

[BL97] J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.

[BM06] G. Bayraksan and D.P. Morton. Assessing solution quality in stochastic programs. *Mathematical Programming*, 108:495–514, 2006.

[CF95] Y. Chen and M. Florian. The nonlinear bilevel programming problem: formulations, regularity and optimality conditions. *Optimization*, 32:193–209, 1995.

[CPS92] R.W. Cottle, J.-S. Pang, and R.E. Stone. *The linear complementarity problem*. Academic Press, Boston, 1992.

[Dem02] S. Dempe. *Foundations of Bilevel Programming*. Kluwer Academic Publishers, 2002.

[EP04] A. Evgrafov and M. Patriksson. On the existence of solutions to stochastic mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 121(1):65–76, 2004.

[Erm88] Y. Ermoliev. Stochastic quasigradient methods. In Y. Ermoliev and R.J.-B. Wets, editors, *Numerical Techniques for Stochastic Optimization*, pages 141–186. Springer, 1988.

[EW88a] Y. Ermoliev and R.J.-B. Wets, editors. *Numerical Techniques for Stochastic Optimization*. Springer Verlag, 1988.

[EW88b] Y. Ermoliev and R.J.-B. Wets. Stochastic programming, an introduction. In Y. Ermoliev and R.J.-B. Wets, editors, *Numerical Techniques for Stochastic Optimization*, pages 1–32. Springer, 1988.

[FL04] R. Fletcher and S. Leyffer. Solving mathematical programs with complementarity constraints as nonlinear programs. *Optimization Methods and Software*, 19(1):15–40, 2004.

[Fle05] M.L. Flegel. *Constraint Qualifications and Stationarity Concepts for Mathematical Programs with Equilibrium Constraints*. PhD thesis, Institute of Applied Mathematics and Statistics, University of Würzburg, 2005.

[FLRS02] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes. Local convergence of SQP methods for mathematical programs with equilibrium constraints. Numerical analysis report na/209, Department of Mathematics, University of Dundee, Dundee, UK, 2002.

[Gai88] A.A. Gaivoronski. Stochastic quasigradient methods and their implementation. In Y. Ermoliev and R.J.-B. Wets, editors, *Numerical Techniques for Stochastic Optimization*, pages 313–351. Springer, 1988.

[Gai04] A.A. Gaivoronski. SQG: stochastic programming software environment. In S.W. Wallace and W.T. Ziemba, editors, *Applications of Stochastic Programming*, MPS-SIAM Series in Optimization, pages 637–670. 2004.

[GR02] C. Greengard and A. Ruszczynski, editors. *Decision Making under Uncertainty. Energy and Power*, volume 128 in The IMA Volumes on Mathematics and its Applications. Springer, 2002.

[Jit84] K. Jittorntrum. Solution point differentiability without strict complementarity in nonlinear programming. *Mathematical Programming Study*, 21:127–138, 1984.

[JR03] H. Jiang and D. Ralph. Extension of quasi-Newton methods to mathematical programs with complementarity constraints. *Computational Optimization and Applications*, 25:123–150, 2003.

[KO04] M. Kočvara and J.V. Outrata. Optimization problems with equilibrium constraints and their numerical solution. *Mathematical Programming, Ser. B*, 101:119–149, 2004.

[LCFar] G.-H. Lin, X. Chen, and M. Fukushima. Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization. *Mathematical Programming*, To appear.

[LF06] G.-H. Lin and M. Fukushima. New reformulations for stochastic nonlinear complementarity problems. *Optimization Methods and Software*, 21(4):551–564, 2006.

[LPR96] Z.Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.

[LS04] X. Liu and J. Sun. Generalized stationary points and an interior-point method for mathematical programs with equilibrium constraints. *Mathematical Programming, Ser. B*, 101:231–261, 2004.

[Mir99] J. Mirrlees. The theory of moral hazard and unobservable behaviour: Part I. *The Review of Economic Studies*, 66(1):3–21, 1999.

[NV77] E.A. Nurminski and P.I. Verchenko. Convergence of algorithms for finding saddle points. *Kibernetika*, 3:112–116, 1977.

[PW99] M. Patriksson and L. Wynter. Stochastic mathematical programs with equilibrium constraints. *Operations Research Letters*, 25:159–167, 1999.

[Sha06] A. Shapiro. Stochastic programming with equilibrium constraints. *Journal of Optimization Theory and Applications*, 128(1):223–243, 2006.

[SR03] A. Shapiro and A. Ruszczynski. *Stochastic Programming. Handbooks in Operations Research and Management Science*, volume 10. Elsevier, 2003.

[SS00] H. Scheel and S. Scholtes. Mathematical programs with equilibrium constraints: stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.

[SX05] A. Shapiro and H. Xu. Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. Preprint, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia. E-print available at: `http://www.optimization-online.org`, 2005.

[Wet89] R.J.-B. Wets. Stochastic optimization. In G.L. Nemhauser, A.H.G. Rinnooy Kan, and M.J. Todd, editors, *Optimization*. North-Holland, 1989.

[WG05] A.S. Werner and A.A. Gaivoronski. Utilization of stochastic programming methods in the analysis of agency problems. In *Bilevel stochastic programming problems: Analysis and application to telecommunications. (Phd Thesis)*, pages 99–131, 2005.

[WW07] Q. Wang and A.S. Werner. Telecom service resale under uncertainty: Profit and consumer surplus implications. In preparation, 2007.

[Wyn01] L. Wynter. Stochastic bilevel programs. In C. Floudas and P. Pardalos, editors, *Encyclopedia of Optimization*. Kluwer, 2001.

[Xu06] H. Xu. An implicit programming approach for a class of stochastic mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 16(3):670–696, 2006.

[Ye05] J.J. Ye. Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 307:350–369, 2005.

[ZL01] J.Z. Zhang and G. Liu. A new extreme point algorithm and its application in PSQP algorithms for solving mathematical programs with linear complementarity constraints. *Journal of Global Optimization*, 14:345–361, 2001.