THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Nonlinear Programming – Robust Models and Applications

Christoffer Cromvik

Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
Gothenburg, Sweden 2009

Nonlinear Programming – Robust Models and Applications
Christoffer Cromvik
ISBN 978-91-7385-315-6

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 2996
ISSN 0346-718X

Department of Mathematical Sciences
Division of Mathematics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31 772 1000

Printed at the Department of Mathematical Sciences
Göteborg, Sweden 2009

Nonlinear Programming – Robust Models and Applications

Christoffer Cromvik

Department of Mathematical Sciences
Chalmers University of Technology
and University of Gothenburg

Abstract

The major theme of this thesis is nonlinear programming with an emphasis on applications and robust models. The thesis has two parts. The first three papers comprise the first part. Here, we discuss robustness properties of optimal solutions to a variety of models. The first two papers concern optimization models known as Stochastic Mathematical Programs with Equilibrium Constraints (SMPEC). These are stochastic optimization problems that have two levels of "decisions": a lower-level one and an upper-level one. The lower-level problem is in the form of a variational inequality, and the upper-level objective function is either the expected value of an objective or the Conditional Value-at-Risk (CVaR). We also consider multiple objective extensions of the SMPEC framework. The stability of optimal solutions due to changes in the underlying probability distribution is analyzed. We also present two applications together with numerical examples: Intensity Modulated Radiation Therapy (IMRT) and traffic network design.

In the third paper, we consider a nonlinear program with multiple objectives which are subjected to uncertainty in the variables and in the parameters. Here we do not use a stochastic programming approach, but instead we wish to analyze robustness as a post-process. Given a particular decision maker, we use his or her preferences to assess the robustness of optimal solutions. This is accomplished through the construction of a utility function which reduces the multi-objective problem into a single-objective problem.

The second part of the thesis, corresponding to the fourth paper, concerns the problem of numerically folding an airbag. We approximate the

airbag geometry by a quasi-cylindrical polyhedron, and we show how Origami mathematics can be used to derive a folding pattern that will collapse the polyhedron. The actual folding problem is solved through the formulation of a nonlinear program whose optimal solution corresponds to the coordinates of the vertices of the flattened polyhedron. The method is demonstrated on a computer model of a passenger airbag.

Keywords: optimization, robustness, multi-objective optimization, stochastic mathematical programs with equilibrium constraints, conditional value-at-risk, IMRT, traffic network design, airbag folding, Origami

Dissertation

This thesis consists of a short review and four papers:

Paper I: Cromvik, C. and Patriksson, M. *On the robustness of global optima and stationary solutions to stochastic mathematical programs with equilibrium constraints, part I: Theory.* Journal of Optimization Theory and Applications, 2010 144(3) (to appear)

Paper II: Cromvik, C. and Patriksson, M. *On the robustness of global optima and stationary solutions to stochastic mathematical programs with equilibrium constraints, part II: Applications.* Journal of Optimization Theory and Applications, 2010, 144(3) (to appear)

Paper III: Cromvik, C. and Lindroth, P. *Robust multi-objective optimization based on a user perspective.* Preprint 2009:31, Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg (submitted to Annals of Operations Research)

Paper IV: Cromvik, C. and Eriksson, K. *Airbag folding based on origami mathematics.* Proceedings on The Fourth International Conference on Origami in Science, Mathematics, and Education (4OSME). A. K. Peters, 2009 (to appear)

Acknowledgments

First and foremost, I would like to thank Michael Patriksson. Michael has been my supervisor since my Licentiate degree, and he has been very helpful, inspirational and encouraging.

I am also very grateful for the collaboration on biological optimization of treatment plans in IMRT that Michael and I have had together with Caroline Olsson, Karl-Axel Johansson, Anna Bäck and Niclas Petersson at the Department of Radiophysics at Sahlgrenska University Hospital. In particular, I am very grateful to Caroline for learning me a lot about radiobiology and IMRT and for the many interesting discussions that we have had.

I would like to thank the advisors of my Licentiate project, Stig Larsson at Chalmers, Kenneth Eriksson at University West and Bengt Pipkorn at Autoliv Research, for their all their help and support and for giving me the opportunity to become a Ph. D. student. I also wish to thank Krystoffer Mroz and Magnus Eriksson at Autoliv Research for their help with airbag folding.

A special thanks goes to my co-author and friend Peter Lindroth. I think it was a nice initiative of us to write an article together, and I have enjoyed it a lot.

Autoliv Research has financially supported my Licentiate project and I am very grateful for this. Gothenburg Mathematical Modeling Centre (GMMC) is also greatly acknowledged for financially supporting several of my travels to conferences.

Many thanks to my colleagues and great friends at the Department of Mathematical Sciences for making it enjoyable to work. In particular I would like to mention Karin Kraft, David Heintz, Thomas Ericsson, Anna Nyström, Mohammad Asadzadeh, Adam Wojciechowski, Ann-Brith Strömberg, Karin Thörnblad, Niklas Ericsson, Fardin Saedpanah, Ali Mesforush, Niklas Eriksen, Göran Starius and Fredrik Lindgren.

Finally, I would like to thank my parents, my sister with family, and especially my wife Julia and my daughter Emelie for their great support.

Christoffer Cromvik
Göteborg, August 2009

Contents

1 Introduction

The purpose of the following sections is to introduce the topics of the papers of the thesis. The material is highly condensed, and we only consider those topics that arise in the papers. That is to say, the following sections are not a complete introduction to nonlinear optimization.

The introduction is organized in the following way. The first section of continuous nonlinear programming provides some theory and algorithms. After that, we summarize and comment each of the appended papers. These sections also serve as introductions to the papers with some background information.

2 Nonlinear programming: An introduction

Already in the 17th and 18th century, some aspects of mathematical programming were considered with the works of Bernoulli, Leibnitz, Euler and Lagrange, among others. In particular, during this time, Euler and Lagrange developed the theory of calculus of variations. Earlier, Johann and Jacob Bernoulli posed the Brachistochrone problem, which can be solved by calculus of variations. It can be formulated as follows: A bead slides from point A to point B in a vertical plane; find the optimal shape of the curve for the bead such that it reaches point B in minimum time assuming that it starts from rest and is accelerated only by gravity.

In general, a mathematical program posed in a finite dimensional space can often be formulated with a vector of decision variables $x \in \mathbb{R}^n$ whose numerical values are constrained to some set $S \subseteq \mathbb{R}^n$, and with a function $f : \mathbb{R}^n \to \mathbb{R}$ which takes the value of x as input and which should be minimized, i.e.,

$$\underset{x \in S}{\text{minimize}}\, f(x).$$

The problem should be interpreted in the following way: we seek a feasible vector $x \in S$ such that f attains its minimum (optimal) value over its domain. Such a vector is denoted the optimal solution. It is common to describe S by the set $\{\, x \mid g_i(x) \leq 0,\ i = 1, \ldots, m_g, h_j(x) = 0,\ j = 1, \ldots, m_h \,\}$, where g_i and h_j are constraint functions of the same type as f. In this case the optimization program reads

$$\text{(NLP)} \quad \begin{cases} \underset{x}{\text{minimize}} & f(x), \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m_g, \\ & h_j(x) = 0, \quad j = 1, \ldots, m_h. \end{cases}$$

If the objective function f or any of the constraint functions g_i, h_j, are nonlinear, the optimization problem is a nonlinear program.

The formulation (NLP) is not as general that it includes all nonlinear programs. For example, mathematical programs with complementarity constraints (MPCC), that are special cases of the problems studied in Paper I, are formulated as

$$\text{(MPCC)} \quad \begin{cases} \text{minimize} & f(x), \\ \text{subject to} & g(x) \leq 0^{m_g}, \\ & 0^p \leq r(x) \perp s(x) \geq 0^p, \end{cases}$$

where $a \perp b \Leftrightarrow a^{\mathrm{T}} b = 0$, for vectors a and b of the same dimensions.

Since the goal of optimization is to find the optimal solution, the notion and theory of optimality is essential. We distinguish between two types of optimality: local and global optimality. A feasible solution x^* is a *globally* optimal solution if

$$f(x^*) \leq f(x), \quad \forall x \in S.$$

A feasible solution x^* is a *locally* optimal solution if there exists an $\varepsilon > 0$ such that

$$f(x^*) \leq f(x), \quad \forall x \in S \cap B_\varepsilon(x^*),$$

where $B_\varepsilon(x^*) = \{x \mid \|x - x^*\| < \varepsilon\}$.

The theory of Karush, Kuhn and Tucker (KKT) provides a convenient way of characterizing optimality for continuously differentiable nonlinear programs. We use a non-smooth version of the KKT conditions in Paper I to derive robustness for optimal solutions. To use the KKT as optimality conditions, it is required that the constraint set S has a certain regularity. A *constraint qualifications* (CQ) is used to verify this regularity. There are several constraint qualifications, and depending on the formulation of the problem, one may be easier to check than the other. For example, the requirement of linear independence of the constraint gradients is perhaps the most immediate CQ. Consider (NLP) and let $A(x) = \{i \mid g_i(\bar{x}) = 0\}$ denote the active set of the inequality constraints at some feasible solution $\bar{x}$. The *linear independence constraint qualification* (LICQ) holds at $\bar{x}$ if the vectors

$$\{\, \nabla g_i(x^*) \mid i \in A(x^*) \,\} \cup \{\, \nabla h_j(x^*) \mid j = 1, \ldots, m_h \,\},$$

are linearly independent. Another CQ is the Mangasarian–Fromovitz constraint qualification, which is used in Paper I. It holds at $\bar{x}$ if the vectors

$$\{\nabla h_j(\bar{x}) \mid j = 1, \ldots, m_h\}$$

are linearly independent, and there exists a vector $d \in \mathbb{R}^n$ such that

$$\nabla h_j(\bar{x})^{\mathrm{T}} d = 0, \ j = 1, \ldots, m_h, \qquad \nabla g_i(\bar{x})^{\mathrm{T}} d < 0, \ i \in A(\bar{x}).$$

Naturally, if LICQ holds for a solution $\bar{x}$, then MFCQ also holds for that solution $\bar{x}$. The most general constraint qualification is the Abadie CQ. (All other CQs imply that the Abadie CQ holds.) Bazaraa et al [BSS06] provide a detailed list of constraint qualifications and their relations.

A feasible solution $\bar{x}$ to (NLP) is referred to as a KKT-point if there exist multipliers $\mu \in \mathbb{R}_+^{m_g}$ and $\lambda \in \mathbb{R}^{m_h}$ such that $\mu_i g_i(\bar{x}) = 0$ for $i = 1, \ldots, m_g$ and

$$\nabla f(\bar{x}) + \sum_{i=1}^{m_g} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{m_h} \lambda_j \nabla h_j(\bar{x}) = 0^n. \tag{1}$$

If a constraint qualification such as LICQ or MFCQ holds, then each local optimal solution is a KKT-point ([BSS06, Theorem 4.37]). If the functions f and g_i, $i = 1, \ldots, m_g$, are convex and h_j, $j = 1, \ldots, m_h$, are affine, then a KKT-point is a global optimal solution ([BSS06, Theorem 4.38]).

If we introduce the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^{m_g} \times \mathbb{R}^{m_h} \to \mathbb{R}$ with

$$L(x, \mu, \lambda) = f(x) + \sum_{i=1}^{m_g} \mu_i g_i(x) + \sum_{j=1}^{m_h} \lambda_j h_j(x),$$

we see that the KKT-condition (1) is equivalent to stationarity of the Lagrangian in terms of x. With this in mind, we often use the term *stationary point* for a KKT-point.

We would like to point out that optimality conditions are not only used for checking optimality. Many optimization algorithms, such as the sequential quadratic programming algorithm in Section 2.2, are in fact based directly on these conditions.

Due to the nature of optimization problems, most problems need to be solved numerically with an iterative scheme. We will concentrate on two numerical methods that are used in the articles in this thesis: the augmented Lagrangian method and sequential quadratic programming. They are both applicable to nonlinear programs with continuously differentiable objective and constraint functions. There are of course many other popular numerical methods. For example, interior-point methods ([FGW02]) for nonlinear programs have gained substantial interest over the last years.

2.1 Augmented Lagrangian method

The augmented Lagrangian method is an exterior penalty method which means that we penalize infeasibility. If all constraints are penalized, the nonlinear program reduces to an unconstrained problem; however, in some cases it may be advantageous to only penalize some of the constraints. The penalty is controlled by a penalty parameter and at each major iteration,

a subproblem is solved. By increasing the penalty successively between the major iterations, the iterates are forced towards feasibility in the end. We consider the problem (NLP) without inequality constraints and with a general closed constraint set $X \subseteq \mathbb{R}^n$ that is not penalized:

$$\begin{aligned} &\underset{x \in X}{\text{minimize}} \quad f(x), \\ &\text{subject to} \;\; h_j(x) = 0, \quad j = 1, \ldots, m_h. \end{aligned} \tag{2}$$

(Note that any inequality constraints can be included in (NLP) with slack variables and with simple bounds added to X.) Let

$$\begin{aligned} h(x) &:= (h_1(x), \ldots, h_{m_h}(x))^{\mathrm{T}}), \\ \nabla h(x) &:= (\nabla h_1(x), \ldots, \nabla h_{m_h}(x)). \end{aligned}$$

We define the augmented Lagrangian function $L_c : \mathbb{R}^n \times \mathbb{R}^{m_h} \to \mathbb{R}$ by

$$L_c(x, \lambda) = f(x) + \lambda^{\mathrm{T}} h(x) + \tfrac{c}{2} \|h(x)\|^2,$$

where $c > 0$ is a penalty parameter and λ is the multiplier vector. The augmented Lagrangian is identical to the Lagrangian for the problem

$$\begin{aligned} &\underset{x \in X}{\text{minimize}} \quad f(x) + \tfrac{c}{2} \|h(x)\|^2, \\ &\text{subject to} \;\; h_j(x) = 0, \quad j = 1, \ldots, m_h. \end{aligned}$$

This problem has the same (global) optimal solution as problem (2), but the objective function has been augmented with a penalty term. Incidentally, this form reveals that there is a connection between the augmented Lagrangian method and the dual method known as the proximal minimization algorithm ([BT89]). The idea of the augmented Lagrangian method is to solve a sequence of problems of the form

$$\begin{aligned} &\text{minimize} \quad L_{c^k}(x, \lambda^k), \\ &\text{subject to} \;\; x \in X, \end{aligned} \tag{3}$$

where $\{\lambda^k\}$ is a bounded sequence and $\{c^k\}$ is a positive increasing sequence of penalty parameters with $c^k \to \infty$. The optimal solution x^k to (3) is used as a starting point for the next problem with λ^{k+1} and c^{k+1}. It can be shown that globally optimal solutions to the subproblems (3) converge to a global optimal solution to (2) ([Ber82, Prop 2.1]) under weak assumptions. Convergence can also be proven for local optimal solutions as well as near-stationary solutions ([Ber82, Prop 2.2, 2.3]).

If we choose $\lambda = 0^{m_h}$, the augmented Lagrangian method reduces to the standard exterior quadratic penalty method. This method may suffer from ill-conditioning, which the partial Hessian reveals:

$$\nabla^2_{xx} L_c(x, 0^{m_h}) = \nabla^2 f(x) + c\nabla h(x)\nabla h(x)^{\mathrm{T}} + c\sum_{i=1}^{m_h} h_i(x)\nabla^2 h_i(x).$$

If the value of c is large and x is almost feasible, the middle term dominates, and the ill-conditioning arises since it is a rank m_h matrix (at best).

By properly choosing λ^k, this ill-conditioning can be avoided. From the expression of the partial gradient of the augmented Lagrangian,

$$\nabla_x L_{c^k}(x, \lambda^k) = \nabla f(x) + \nabla h(x)(\lambda^k + ch(x)),$$

and the KKT-conditions with Lagrange multipliers λ^*, we see that λ^k should be chosen as

$$\lambda^* \approx \lambda + c^k h(x^k).$$

It can be shown ([Ber82]) that if λ is updated with

$$\lambda^{k+1} = \lambda^k + c^k h(x^k),$$

then there is a threshold value of c^k which gives local convergence, and also that the convergence is accelerated from linear to superlinear speed if also $c^k \to \infty$.

The augmented Lagrangian method is the basis of the optimization software LANCELOT [GOT04]. It uses a gradient projection algorithm (see e.g. [NW06]) for solving the possibly bound-constrained subproblems (3). The software LANCELOT is used in Paper II to solve the problems in intensity modulated radiation therapy. LANCELOT was chosen mainly because it is written in Fortran 90, as is the subroutines for the application, and because it can solve large-scale problems in a reasonable time.

2.2 Sequential Quadratic Programming

Sequential quadratic program (SQP) is a popular method for small and medium-scale nonlinear programs (NLP). The method proceeds by solving a sequence of quadratic problems that are designed to compute search directions. SQP can be seen as applying Newton's method to solve the KKT-conditions (1) as a system of equations. Let x^k be the current iterate, and d^k the next search direction. The quadratic subproblems are of the following form

$$\begin{aligned} \underset{d^k}{\text{minimize}} \quad & \tfrac{1}{2}(d^k)^{\mathrm{T}} B^k d^k + (d^k)^{\mathrm{T}} \nabla f(x^k), \\ \text{subject to} \quad & \nabla g(x^k)^{\mathrm{T}} d + g(x^k) \leq 0^{m_g}, \\ & \nabla h(x^k)^{\mathrm{T}} d + h(x^k) = 0^{m_h}, \end{aligned} \tag{4}$$

where B^k is the Hessian of the Lagrangian or an approximation of it. Consider the KKT-conditions for the problem (NLP) without the inequality constraints as a system of equations

$$F(x,\lambda) = \begin{pmatrix} \nabla f(x) - \nabla h(x)\lambda \\ h(x) \end{pmatrix} = 0^{n+m_h}.$$

If we were to apply Newton's method at the current iterate (x^k, λ^k), we get

$$\begin{pmatrix} \nabla^2_{xx} L(x^k,\lambda^k) & \nabla h(x^k) \\ \nabla h(x^k)^{\mathrm{T}} & 0 \end{pmatrix} \begin{pmatrix} d^k \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ -h(x^k) \end{pmatrix}, \tag{5}$$

where $x^{k+1} = x^k + d^k$. If we assume that the constraint Jacobian $\nabla h(x^k)$ has full row rank, and that the Hessian of the Lagrangian is positive definite on the nullspace of the linearized constraints, then the solution d^k and λ^{k+1} to equation (5) is equal to the optimal solution and the optimal Lagrange multipliers respectively, to problem (4) provided B^k is chosen as the Hessian of the Lagrangian.

Local convergence of the SQP method can be established by assuming that LICQ and strict complementarity hold as well as a second-order requirement on the Hessian of the Lagrangian (see [BT95] for details).

To expand the radius of converges and to stabilize the algorithm, there is an enhancement called globalization strategy that can be used. One such is to use a merit function. The merit function measures progress in each iteration by incorporating both the objective value and the level of infeasibility. One popular choice is the non-differentiable function

$$\phi_\rho(x) = f(x) + \rho \left(\sum_{i=1}^{m_g} \| \max\{0, g_i(x)\} \|_1 + \sum_{j=1}^{m_h} \|h_j\|_1 \right), \tag{6}$$

where $\|\cdot\|_1$ denote the l_1-norm and ρ is a penalty parameter. This function is exact in the sense that, under a second-order requirement on the Hessian, there is a threshold value for ρ for which a local minimum to (NLP) is a local minimizer of ϕ_ρ ([BSS06, Theorem 9.3.1]). Instead of taking a step with unit step length $x^{k+1} = x^k + d^k$, we perform a line search using backtracking. Let $\alpha^l = 1/2^l$, for iterations $l = 1, 2, \ldots$, and let $\sigma < 1$ be a small number. Typically, the step length α^l is accepted, i.e., $x^{k+1} = x^k + \alpha^l d^k$, if

$$\phi_\rho(x^k + \alpha^l d^k) \leq \phi_\rho(x^k) + \sigma \alpha^l \mathrm{D}\phi_\rho(x^k; d^k),$$

where $\mathrm{D}\phi_\rho(x^k; d^k)$ denotes the the directional derivative of ϕ_ρ along d^k.

Another strategy is to use a "filter" ([FLT02]). The objective value and the level of infeasibility is stored for all iterates in a filter, and in a line

search, the step length is accepted if the iterate passes the filter. The filter is based on multi-objective optimization. (see Section 2.3). The iterate is said to dominate if both the objective value and the level of infeasibility is less than the other iterates in the filter. A new iterate is accepted if it is not dominated by any other stored iterate.

The optimization software SNOPT [GMS05] implements an SQP method and it is used in Paper II to solve the problem in traffic network design. SNOPT was chosen because it is able to solve large-scale problems in reasonable time and because it is reported to be successful at solving mathematical programs with equilibrium constraints (MPCC) ([FL04, FLRS06]), which is the problem that the modeling of the traffic network design problem results in. The MPCC need to be converted into standard form in order to be solved by SNOPT, and we used the formulation:

$$\begin{aligned} \text{minimize} \quad & f(x), \\ \text{subject to} \quad & g(x) \leq 0^{m_g}, \\ & r(x) \geq 0^{p}, \\ & s(x) \geq 0^{p}, \\ & r(x)^{\mathrm{T}} s(x) \leq 0. \end{aligned}$$

In Paper IV, we use an in-house implemented (in Fortran 90) SQP method. We use the function (6) as a merit function and solve the quadratic programs with a primal-dual interior-point method. The main computational cost, at each iteration, is to solve a large linear system of equations that have a symmetric indefinite matrix. We use the software package PARDISO [SG04, SG06] to solve these equations.

2.3 Multi-objective problems

There are often several objectives that are considered for minimization for a practical optimization problem. The application of radiation therapy in Paper II is an example of this. When faced with several objectives, we may consider a multi-objective optimization problem (MOP), which we present in the following form

$$(\text{MOP}) \begin{cases} \underset{x}{\text{minimize}} & (f_1(x), \ldots, f_k(x)), \\ \text{subject to} & g(x) \leq 0^{m_g}, \end{cases}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, k$ and $g : \mathbb{R}^n \to \mathbb{R}^{m_g}$. To make sense of the minimization of the vector of objective functions, we have to introduce specific notions of "optimality". A feasible solution $\bar{x}$ is called a *Pareto optimal* solution if there are no feasible solutions x such that $f_i(x) \leq f_i(\bar{x})$

for $i = 1, \ldots k$ with $f_j(\bar{x}) < f_j(x)$ for at least one j. The set of all Pareto optimal solutions is called the *Pareto optimal set* $\mathcal{P}$. The corresponding objective values form a "front" which is called the *Pareto optimal front* $f(\mathcal{P})$. Furthermore, a feasible solution $\bar{x}$ is called a *weakly Pareto optimal* solution if there are no feasible solutions x such that $f_i(x) < f_i(\bar{x})$ for $i = 1, \ldots k$.

Example 2.1 *Consider the problem to* minimize $(x^2, (x-1)^2)$ *for* $x \in [-1, 2]$. *The Pareto optimal set is* $\mathcal{P} = [0, 1]$, *and the Pareto optimal front is shown in Figure 1.*

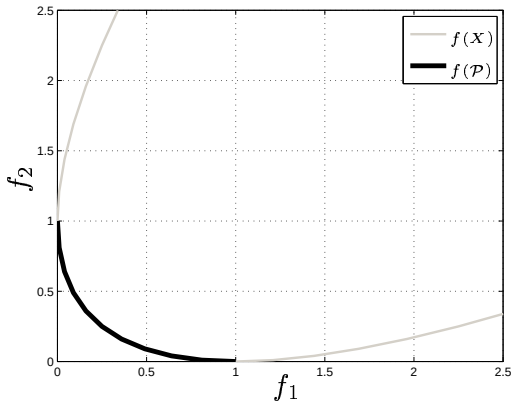


Figure 1: The objective values $f(X)$ and the Pareto front $f(\mathcal{P})$ for Example 2.1.

Solution methods for multi-objective problems are designed to find the Pareto optimal set, or at least a representation of it. We can distinguish between two categories of methods: one targets the multi-objective problem directly and iteratively finds presumed Pareto points. In the other category, the multi-objective problems are reformulated as single-objective problems.

The genetic algorithm NSGA-II [DAPM00] is an example of the first category. Since genetic algorithms compare function values for different generations, it is natural to use dominance for several objective functions instead. The algorithm NSGA-II is used in Paper III. In the second category, we consider two methods: the weighted sum method and the epsilon-constraint method. In the weighted sum method, the single-objective problem is formed from a linear combination of the objective functions

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \sum_{i=1}^{k} w_i f_i(x), \\ \text{subject to} \quad & g(x) \leq 0^{m_g}, \end{aligned} \tag{7}$$

where $w_i > 0$ for $i = 1, \ldots, k$. The idea is to vary the coefficients (or, weights) w which produces a set of Pareto optimal solutions. An advantage of this method is that its implementation is simple. The downside is that only convex Pareto optimal fronts can be resolved ([Mie99]), and it is difficult to find a good representation of the front since the mapping from the weights strongly depends on the functions involved ([DD97]).

In the epsilon-constraint method, all but one objective f_j are placed as additional constraints:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_j(x), \\ \text{subject to} \quad & f_i(x) \leq \varepsilon_i, \quad i = 1, \ldots, k,\ i \neq j, \\ & g(x) \leq 0^{m_g}, \end{aligned} \tag{8}$$

where $\varepsilon \in \mathbb{R}^{k-1}$. We are able to resolve the Pareto front by discretizing each ε_i and solve (8) for all combinations of ε_i. The method has the advantage that the problem of resolving the front is easier than with the weighted sum. A downside is the added complexity with the additional constraints. We use the epsilon-constraint method in Paper II for the problems in intensity modulated radiation therapy.

For further information on multi-objective problems, see, e.g., the monographs [Mie99, Ehr05].

3 Summary of and comments on the papers

The presentation of each paper is subdivided into three sections: background, contributions and future work.

3.1 Paper I: On the robustness of global optima and stationary solutions to stochastic mathematical programs with equilibrium constraints, part 1: Theory

3.1.1 Background

We consider a special type of nonlinear program that is called stochastic mathematical program with equilibrium constraints (SMPEC). Its deterministic version is the MPEC.

The idea of this paper comes from an article by Evgrafov and Patriksson [EP03], in which structural optimization problems, modeled as SMPECs, were studied and a few basic results on robustness were obtained. We first had the idea that we could obtain results on robustness for an application in intensity modulated radiation therapy (see Paper II), but we later realized we could remain in the more general setting of the SMPEC.

MPECs are hierarchical optimization problems, in which the problem comprises two levels of "decisions": a lower-level one and an upper-level one. We also consider two types of variables: decision variables and response variables. The decision variables parameterize the lower-level problem; its solution is the value of the response variables. The lower-level problem is formulated as a variational inequality. The problem has found its use in many applications such as traffic network design [Ral08], structural optimization [EP03, EPP03], chemical engineering [BRB08], etc.

The MPEC is formulated as the problem to

$$\text{(MPEC)} \begin{cases} \underset{(x,y)}{\text{minimize}} & f(x,y), \\ \text{subject to} & x \in X, \\ & y \in C(x), \quad F(x,y)^{\mathrm{T}}(y_0 - y) \geq 0, \qquad \forall y_0 \in C(x), \end{cases}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $y \in \mathbb{R}^m$, $C(x) \subseteq \mathbb{R}^m$ is a closed convex set and $F(x,\cdot) : C(x) \to \mathbb{R}^m$ is smooth. When the solution to the lower-level problem is non-unique for a fixed x, the model should be interpreted as y being chosen such that the objective function is minimized given x.

If we let $C = \mathbb{R}^m_+$, then the variational inequality becomes a complementarity constraint, $0^m \leq F(x,y) \perp y \geq 0^m$. Mathematical programs with complementarity constraints (MPCC) are notoriously difficult since they lack standard constraint qualifications ([SS00]). This suggests that also more general MPECs may be difficult to solve.

For additional general information on variational inequalities we refer to Facchinei and Pang [FP03a, FP03b], and Dontchev and Rockafellar [DR09]. For recent work on the numerical solution of MPECs, we refer to [FLRS06, LLCN06].

As an example of an application of MPEC, we consider the Stackelberg game ([VS52, LPR96]), which is a leader-follower game and an extension of the Nash game. One of its uses is in electric transmission pricing polices [HK92]. We assume that there are m players, called followers, and one leader. Each follower has a strategy $y_i \in Y_i \subseteq \mathbb{R}^{m_i}$ and wishes to minimize its economic cost, given the decisions of the other players. The leader also wishes to minimize its economic cost f by choosing a strategy $x \in X \subseteq \mathbb{R}^n$. Let $\phi_i(x^{\text{given}}, \cdot, y^{\text{given}}_{\neq i}) : Y_i \to \mathbb{R}$ denote the economic cost for follower i and let us assume it is convex and continuously differentiable. We also let $Y_i(x)$ be a closed and convex set. Each player observes the decisions of the other players and react optimally accordingly. A strategy $y^* \in \prod_{j=1}^m Y_j(x)$ is called a Nash equilibrium if $y_i^* \in \operatorname{argmin}\{\phi_i(x^{\text{given}}, y_i, y^{\text{given}}_{\neq i}) \mid y_i \in Y_i(x)\}$ for $i = 1, \ldots, m$.

The leader can anticipate the decisions of the other players (the fol-

lowers) when choosing optimal strategy x. Let $f(x,y)$ denote the leader's economic cost, and let $F_i = \nabla_{y_i}\phi_i(x,y)$. The Stackelberg game problem can be formulated as the following MPEC

$$\begin{aligned} \underset{x,y}{\text{minimize}} \quad & f(x,y), \\ \text{subject to} \quad & x \in X, \\ & y \in \prod_{i=1}^{m} Y_i(x), \quad F(x,y)^{\mathrm{T}}(y_0 - y) \geq 0, \qquad \forall y_0 \in \prod_{i=1}^{m} Y_i(x). \end{aligned}$$

It is natural to consider the variational inequality and the objective function to be subjected to uncertainty. In the Stackelberg game for example, both the leader's and the follower's economic costs may depend on uncertain external parameters. If we wish to find an optimal solution that is best in an average sense, we may consider the SMPEC which was formulated by Patriksson and Wynter [PW99]:

$$(\text{SMPEC}_\Omega) \begin{cases} \underset{(x,y(\cdot))}{\text{minimize}} \ \mathrm{E}_\omega[f(x,y(\omega),\omega)] := \displaystyle\int_\Omega f(x,y(\omega),\omega)\,\mathrm{P}(d\omega), \\ \text{subject to} \ x \in X, \\ \qquad\qquad y \in C(x,\omega), \quad F(x,y,\omega)^{\mathrm{T}}(y_0 - y) \geq 0, \\ \qquad\qquad\qquad\qquad \forall y_0 \in C(x,\omega), \ \mathrm{P\text{-a.s.}} \end{cases}$$

where $y : \Omega \to \mathbb{R}^m$ is a random element of the probability space $(\Omega, \Theta, \mathrm{P})$. In view of stochastic programming with recourse ([BL97]), SMPEC is considered as a here-and-now type of problem, where decisions x should be taken before any realizations of uncertain data. When the solution to the lower-level problem is non-unique for a fixed x and ω, the model should be interpreted as y being chosen such that the objective function is minimized given x and ω.

Consider the two-stage linear program with fixed recourse ([BL97]):

$$\begin{aligned} \text{minimize} \quad & c^{\mathrm{T}}x + \mathrm{E}[\min q(\omega)^{\mathrm{T}} y(\omega)], \\ \text{subject to} \quad & T(\omega)x + W y(\omega) = h(\omega), \\ & Ax = b, \\ & x \geq 0, \\ & y(\omega) \geq 0, \end{aligned}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and q, T and h are random elements of corresponding sizes. A decision x is made in the first stage, and associated with this are vectors b and c and the matrix A. In the second stage, a random

event $\omega \in \Omega$ occurs, and the data $q(\omega)$, $h(\omega)$ and $T(\omega)$ become known. The two-stage linear program is used in several applications, see, e.g., [BL97] and references therein. If we let $C(x,\omega) = \{y \in \mathbb{R}^m_+ \mid Wy = h(\omega) - T(\omega)x\}$, $F(x, y(\omega), \omega) = q(\omega)$, $f(x, y(\omega), \omega) = c^{\mathrm{T}}x + q(\omega)^{\mathrm{T}}y(\omega)$ and $X = \{x \in \mathbb{R}^n_+ \mid Ax = b\}$, then the two-stage linear program can med modeled as an SMPEC.

If the probability distribution is discrete, we may clone the response variables, one for each scenario, and replace the integral in the objective function with a sum. For a continuous probability distribution, we (probably) have to discretize. We consider sample average approximation (SAA), where the idea is to draw N independently and identically-distributed samples ω^k and solve, for increasing N, the deterministic MPEC problem to

$$(\text{SMPEC})^N \begin{cases} \underset{x}{\text{minimize}} \ \hat{f}_N := \dfrac{1}{N}\displaystyle\sum_{k=1}^{N} f(x, y^k, \omega^k), \\ \text{subject to} \quad x \in X, \\ \qquad\qquad y^k \in C(x, \omega^k), \qquad F(x, y^k, \omega^k)^{\mathrm{T}}(y_0^k - y^k) \geq 0, \\ \qquad\qquad\qquad\qquad\qquad \forall y_0^k \in C(x, \omega^k), \ k = 1, \ldots, N. \end{cases}$$

3.1.2 Contributions

Although the formulation of the SMPEC implies that we consider the uncertainty explicitly, we are interested in which effect perturbations in the probability distribution has on the solution. In particular, we are interested in if an optimal solution to the SMPEC changes continuously due to a continuous change in the probability distribution. We show that global optima and stationary solutions to (SMPEC_Ω), where we consider C to be fixed, are robust in this sense. We also analyze two extensions of the SMPEC model: one in which we replace the expected value in the objective function with the risk measure known as Conditional Value-at-Risk (CVaR); in the other we consider the multi-objective SMPEC model. For CVaR we establish robustness for global optima and stationary solutions. For the multi-objective version, we establish robustness of weakly Pareto optimal and weakly Pareto stationary solutions.

We also study the discretization scheme sample average approximation, which is convergent for both global optima and stationary solutions. Together with the results on robustness, we can hence, to some degree, motivate the use of SMPECs in practice.

3.1.3 Future work

The collected research on SMPEC is rather recent, and there are obviously many opportunities for future research. One particular important subject is

numerical algorithms. The successful application of the SMPEC in practice will depend much on the availability and robustness of numerical software. The problem is twofold: the need to discretize for continuous probability distributions and deterministic algorithms for large-scale problems with complementarity constraints (or bilevel optimization).

3.2 Paper II: On the robustness of global optima and stationary solutions to stochastic mathematical programs with equilibrium constraints, part 2: Applications

3.2.1 Background

Our first idea was to add numerical examples to the first paper on SMPEC. The intention was to demonstrate the general use of the SMPEC and also to exemplify the requirements that we state for the problem in order to prove robustness and convergence of the discretization scheme. Due to the subsequent length of the paper and upon recommendations by the associate editor of JOTA, we decided to write a second paper and give more details on the applications in question.

Intensity modulated radiation therapy is a topic which we have worked on since 2007. It has been one of the main motivators for the research in terms of theory and focus. It is an interesting application of nonlinear optimization theory which has received a lot of research interest over the years (see e.g. [Bor06]), but it is also a very important application because results may in the future improve the quality of life for patients. Intensity modulated radiation therapy is an interdisciplinary subject. Radiation therapy involves oncology and radiation physics, and since the technique intensity modulated results in an inverse problem, the topic also involves mathematics. As such, it requires a lot of knowledge and cooperations, both between researchers, but also with clinicians.

Although there have been a lot of theoretical advances on how to improve the treatment, it is important to note that these advances require an accurate delivery of the radiation. The effect of various uncertainties may, unless considered, prevent a successful treatment. For example, with recent equipment and methodology, clinicians are able to shape the doses more accurately in theory. This would reduce the risk of complications in principle, but with a positional uncertainty of the patient, the accuracy may be useless or even worsen the plan, since the doses are expected to change rapidly over the regions.

There has also been a lot of research in using what is known as biological objective functions (see e.g. [Bra95, WMNSU02, CD02, OJW05]) to derive good treatment plans. These objectives are based on radiobiological models

whose parameters are extracted from previous clinical studies. Since the parameters are uncertain, it should be advisable to include this uncertainty into the optimization problem.

3.2.2 Traffic network design

Traffic network design and the traffic assignment problem is a research topic which has been very active since the 1950s. Traffic network design is an interesting application of nonlinear programming; partly because the results are very concrete and partly because almost everyone can relate to some of the areas such as the impact of congestion.

The objective of traffic network design is to change or construct a network, for example the traffic network in an urban environment, such that some travel times decrease or to make travel times more fair for different groups of travelers. The design can be changed by increasing or decreasing the capacity of roads or by setting tolls.

The amount of travel is a result of many individual decisions. Each traveler has several options on which route to use to get from A to B. The choice depends on the amount of traffic and on the distance. The traveler may also wish to change destination depending on traffic conditions.

The problem of computing the amount of travel given the conditions of the network is called the traffic assignment problem. It is assumed that the amount of travel is static and has reached an equilibrium. This is a fair assumption if we consider longer time durations. The network is represented by a graph and the traffic is represented as flows. Wardrop's user equilibrium condition ([War52]) states that for each Origin–Destination pair (OD), the travel cost[1] for all routes utilized must be equal and minimal. This amounts to an equilibrium problem which can be specified with complementarity constraints. Let $\mathcal{C}$ denote the set of OD pairs, and for each OD pair k, let $h_k \in \mathbb{R}^{m_k}$ denote the vector of route flows. Let also c_k denote the vector of route travel costs, and h denote the vector of route flows over all OD pairs. The equilibrium condition then reads

$$0^{m_k} \leq h_k \perp c_k(h) - \pi_k 1^{m_k} \geq 0^{m_k}, \qquad k \in \mathcal{C}.$$

This complementarity constraint implies that there will only be a positive flow if the cost is minimal.

In traffic network design, we assume that we can alter the path travel costs with design parameters x, i.e., we let $c_k = c_k(x, h)$. We increase the cost if we set a toll or decrease capacity, and likewise, decrease the cost if we increase the capacity. The traffic network design problem can be modeled as

[1] Travel cost can be a combination of travel time and any tolls.

an MPEC, where the lower-level problem is the traffic assignment problem, parameterized by x.

It is quite natural to consider the travel times to be uncertain. Although they are static, they may depend on unforeseen external events such as weather changes or road conditions. To predict the amount of travel, it may be useful to include various scenarios. This can be achieved for traffic network design with the SMPEC model. In that case, we consider various scenarios and find, for example, a design which decreases the amount of travel in an average sense using an expected value as the objective function. This problem has been studied in Patriksson [Pat08a, Pat08b]. We may also focus on reducing the travel times for some more extreme scenarios by using a risk measure as the objective.

For further references on traffic equilibrium models, see the monographs [She85, Pat94] and [MP07].

3.2.3 Intensity modulated radiation therapy

In radiation therapy, cancerous tumors are subjected to ionizing radiation. The objective is to eradicate the tumor while sparing the surrounding tissue and organs at risk. There are two types of radiation treatment: external and internal. In external treatment, the radiation is delivered using a linear accelerator and thereafter directed toward the patient. In internal treatment, a radioactive source is placed inside the patient.

External radiation therapy treatment involves several types of equipment, where the linear accelerator is central. Inside the linear accelerator, electrons are accelerated and hit a metal target which creates high-energy photons. The high-energy photons interact with the cells through elastic and inelastic collisions. This creates electrons, which in turn, may collide with the DNA molecules and thus damage the cells. Some of the damage is repairable, and some is unrepairable. Since cancerous cells are reproducing fast, they are more sensitive to radiation than healthy cells. This is the reason why radiation therapy is given in fractions. A patient undergoes treatment typically one session each weekday for six to eight weeks. This gives the healthy cells sufficient time to repair between the sessions.

Besides the linear accelerator, imaging techniques such as Computer Tomography (CT) are used in the treatment to ordinate the doses. Special regions of interest (ROI) are marked by an oncologist. The tumor and closely located lymph sites are denoted target volume. The Planned Target Volume (PTV) encompass the target volume with an additional margin. Sensitive organs that surrounds the target are denoted Organs At Risk (OAR).

The treatment is organized using a computer software which is called the treatment planning system. The software can visualize the CT scans

and the regions of interest along with the doses. It can also generate dose-volume histograms and other plots to help determining which treatment is best suited. We have used the freely available software CERR [DBC03], see Figure 2.

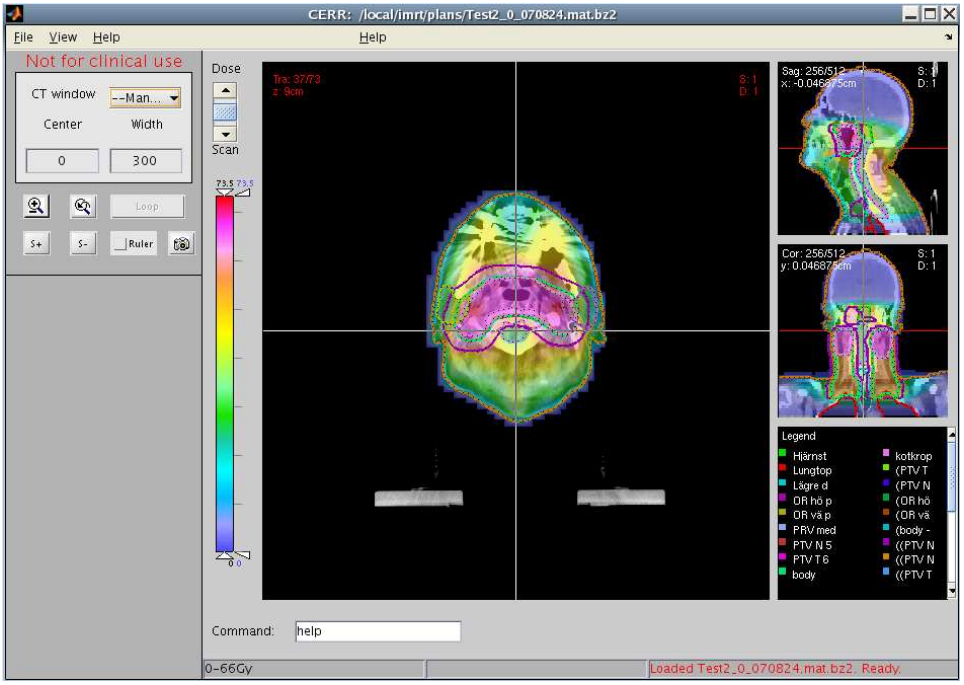


Figure 2: CERR: A computational environment for radiotherapy research. The main window shows a transverse CT scan and isodoses for a treatment plan in the head and neck region.

Intensity Modulated Radiation Therapy (IMRT) is a technique to modulate the intensity such that conformal radiation can be delivered. The modulation is accomplished by using what is known as jaws and Multi-Leaf Collimators (MLC). The modulation is indirect: the linear accelerator delivers radiation continuously, and the jaws and MLCs move on trajectories in order to block certain regions of the beam cross-section. Finding a good treatment plan in IMRT is an inverse problem: we know the doses we want in the ROI, and the problem is to determine which intensities or trajectories can make this happen. The ideal doses are often not attainable. It is often assumed that the doses scale linearly with the intensities in the beam and that the doses are additive. We assume that the beam is discretized with rectangular cells, known as beamlets, and that the ROI are discretized with volume elements, known as voxels. Let $x \in \mathbb{R}^n$ denote the intensities of

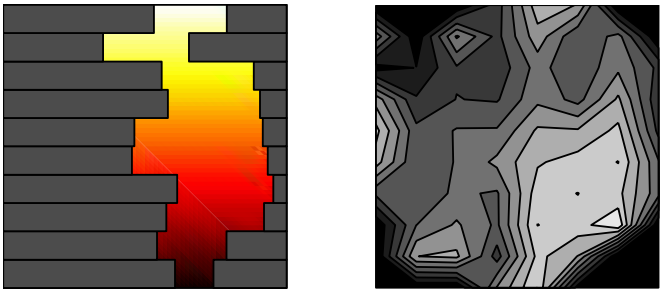


Figure 3: A snapshot of the motion of the multi-leaf collimators (left figure) for the fluence (intensity) profile shown in the right figure for a cross-section of the beam.

the beamlets and let $d \in \mathbb{R}^m$ denote the dose in the voxels. We have the following relation between the intensities and the dose:

$$d = Kx, \tag{9a}$$

$$x \in X, \tag{9b}$$

where K is called the influence matrix and the set X is described by simple bounds. This formulation assumes that we can manipulate the intensities independently.

Another approach is to consider a parameterization of the Multi-Leaf Collimators. There are two types of methods of the motion for the collimators: static and dynamic motion. The idea with the static method is to generate a few segments (shapes) with the collimators, and let each segment be open for a certain time. With the dynamic method, the idea is to compute leaf trajectories and let the beam be open until the collimators stop. To realize either of the methods require that we solve an optimization problem.

Regardless if equations (9) are used or if we use a direct parameterization of the MLCs, we can utilize the MPEC framework to derive a suitable (optimal) treatment plan.

Objective functions in IMRT are categorized as either being physical or biological. The main difference is that a biological objective function uses a biological model to predict the outcome, whereas a physical objective function is a function of the dose alone.

Several objective functions are used in combination to construct an optimization program for a treatment plan. The target and each organ at risk may require an individual objective, and in the end, the oncologist determines if the result is good. Which objective function that should be used

and selecting the parameter values are often determined in an iterative process by examining the end result. This is a time-consuming, but important, process.

Uncertainties arise in various aspects of IMRT: patient motion uncertainty during treatment, patient setup error between sessions, biological uncertainties in the objective functions, machine uncertainty etc. Any information on the uncertainties are natural to include in the optimization problem. The PTV, for example, is used to make certain that the tumor receives a sufficient dose despite of any (modest) motion or position changes. The downside with the approach is that tissue and organs that are close to the tumor and thus overlaps the PTV receive a high dose. If only the extreme positions of the organs are known during movement, then the PTV is the only choice of plan.

If MPEC is used as the modeling framework, it is natural to consider the SMPEC whenever uncertainties are considered. This obviously requires that probability distributions for the parameters are known. If the distributions are unknown or if a conservative plan is required, then Robust optimization [BTN02, CBT06] may be an option.

For more information on IMRT, see e.g. the monograph [Web01] and [LÖ0, Car08].

3.2.4 Contributions

The main contribution of the article is to exemplify the use of SMPEC using numerical examples. In particular, for the traffic network design example, we introduced the CVaR objective as a viable approach.

For IMRT, we considered two examples with uncertainty in the position and uncertainty in the delivery of radiation. Uncertainty in position has been studied by several authors [CBT06, OW06, CZHS05] in various settings, although none using SMPEC. Incorporating uncertainty in the delivery appears to be new and may be interesting to consider in a further, more detailed, study.

3.2.5 Future work

The traffic example presented in the paper is very basic in its deterministic version and so it is natural to consider more realistic networks and more advanced models such as considering elastic demands for which the theory also applies. It would also be interesting to study multiclass models and equity measures in a realistic setting.

For the application of IMRT, it would be interesting to consider biological uncertainty. However, this requires good statistical information on the

response of organs due to inhomogeneous radiation. The information that we have used for position and machine uncertainty have not been based on real statistical data, but in comparison to biological uncertainty, it is easy to consider artificial probabilities and still be able to draw some conclusions. With biological uncertainty on the other hand, it is much harder to draw any conclusions using invalid probabilities.

Finally, we have only used local optimization algorithms, and it would definitely be interesting to consider global ones since not all models we consider are convex.

3.3 Paper III: Robust multi-objective optimization based on a user perspective

3.3.1 Background

The work on *Robust multi-objective optimization based on a user perspective* began with a conversation between the authors on robust multi-objective optimization. The discussion concerned how robustness for multiple objectives should be characterized from a user's perspective. The discussion continued with a collaboration, with the goal to write an article.

The theory of robustness for multi-objective problems in Paper I deals only with sequences of Pareto optimal points. It may be of importance to investigate where on the Pareto surface the solutions in the sequence are located. For example, consider a bicriterion problem with a convex Pareto front and a specific Pareto optimal solution on this front. We consider that a decision maker has chosen this solution for a specific reason. It could be that he or she has considered that the local trade-offs are perfect at this point. This means that the price to pay for improving one objective is too high with respect to the change in the other objective. Consider now that there are uncertainties in the underlying problem and that these may contribute to a change in the objective values. Certainly, it is important how much the objectives changes, but the decision maker would argue that it is also important if the solution changes character, i.e., if the local trade-offs change. This example highlights the main background for this paper. We consider robustness from a decision maker's perspective by introducing a single objective function, called the utility function, which is designed such that the solution that a decision maker chooses (perhaps on the Pareto surface) is optimal. With this single objective function and traditional measures of robustness for single-objective problems, we quantify robustness properties for multi-objective problems.

3.3.2 Contributions

The main contribution is the introduction of the single objective, the utility function, into the measure of robustness for multi-objective problems. This is a novel approach. As a biproduct, we also introduce an alternative characterization of proper Pareto fronts that is more strict than the definition by Geoffrion [Geo68].

3.3.3 Future work

Since the paper is an introductory work, there are lots of possibilities for enhancements and further work. In particular, other definitions of robustness for single-objective problems could be elaborated. We may consider just measuring the expected value of the utility function as a measure, or consider a worst-case scenario which could lead to a minimax formulation.

It would also be interesting to see if it is possible to search for robust solutions simultaneously as the Pareto surface is obtained. For example, if the problem is convex and the weighted sum method is used, then the utility function is already defined for the optimal solution.

3.4 Paper IV: Airbag folding based on origami mathematics

3.4.1 Background

This work was initiated by Bengt Pipkorn at Autoliv Research and began as a master's thesis project. Autoliv Research then financially supported a continuation as a licentiate project.

The problem, in short, was that the methods and software available for numerically folding passenger airbags were not satisfactory, and there was an idea that the mathematical theory of Origami could be used to solve the problem.

Autoliv Research frequently use numerical software to evaluate the performance of airbags. Although real tests are essential, computer simulations have the advantage that many scenarios can be modeled in a short time and at a low cost. Evaluating the performance of airbags requires accurate computer models of the airbag and its surroundings together with accurate numerical methods for simulating elasticity and gas dynamics. Passenger airbags are often tested in an out-of-position situation, which means that in a crash, the passenger hits an expanding, non-inflated, airbag. A computer simulation of this scenario requires a computer model of the folded airbag and here arises the problem of airbag folding: computer models of passenger airbags are only available in its inflated (design) state and passenger airbags have a rather complex three-dimensional geometry, see Figure 4.

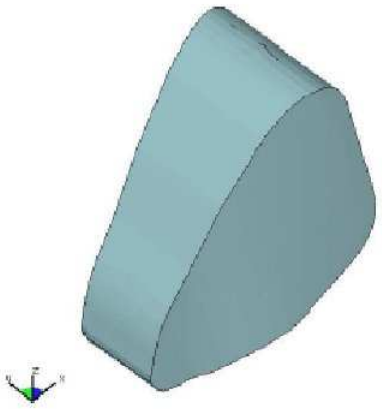


Figure 4: CAD model of a passenger airbag.

Transforming the airbag from its design state to its flat folded state is the subject of numerical airbag folding. Figure 5 shows a real passenger airbag and its folded flat state.

Origami is the ancient art of paperfolding, and besides being an art form, it has found its use in several applications such as the design of a stent in medicine [KTY+06] and for folding large telescopes for transportation [Mye69].

Origami instructions can either be a sequence of figures describing, step by step, how to fold a piece of paper or it can be a crease pattern which consists of lines at the positions of the creases in the final folded model. The situation with folding an airbag is slightly different, since we start with a three-dimensional object, and wish to flatten it. This is accomplished by approximating the airbag by a polyhedron[2], designing a crease pattern on each face, and finally folding the polyhedron according to the crease patterns. This means that we consider each face as a piece of paper; using Origami, we wish to fold it into an object that is still flat (in another, orthogonal, direction) and whose creases coincides with the creases from the crease patterns to adjacent faces.

As an example, Figure 6 shows a model of a paper bag that is often found at supermarkets. It is a rectangular box minus one face, and its crease pattern consists of 21 lines, including the edges of the bag. At each vertex, four lines meet, except at the open top. The creases are either mountain folds or valley folds, and the difference between the number of mountain folds

[2]That is, a not necessarily convex, but always a closed solid in $\mathbb{R}^3$, bounded by plane faces.

Figure 5: "Inflated" and flat folded passenger airbag.

and valley folds are always 2 for an interior vertex ([Lan03]). For example, consider vertex v_1 in Figure 6. Either there are three mountain folds and one valley fold or one mountain fold and three valley folds, depending on in which direction we fold the bag.

3.4.2 Contributions

We represent the airbag by a polyhedron and show how to design a crease pattern for certain types of polyhedra. We provide proofs that there is a folded flat state using the generated creases.

We also describe a procedure to actually fold the polyhedron along the creases. This is achieved by solving an optimization problem whose optimal solution corresponds to the folded flat state.

To illustrate the idea of the optimization problem, we consider folding an object consisting of two triangular sheets and one crease, see Figure 7.

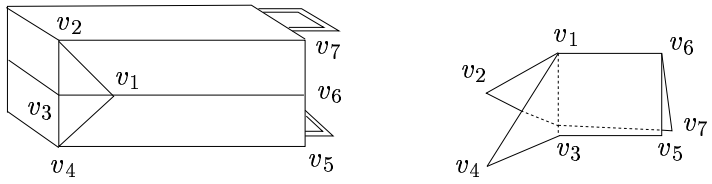


Figure 6: To the left: a paper bag with a corresponding crease pattern. To the right: the corresponding fold of the front face with three mountain folds and one valley fold.

We assume that the vertices $q^1 = (0,0,0)$, $q^2 = (1,0,0)$, $q^3 = (0,1,0)$ are fixed and we denote the fourth vertex by x. Let n^1 and n^2 denote normals to the triangles $\{q^1, q^2, q^3\}$ and $\{q^1, q^3, x\}$ respectively. Consider the following optimization problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & n^1 \cdot n^2, \\ \text{subject to} \quad & \|x - q^1\|^2 = 1, \\ & \|x - q^3\|^2 = 2, \\ & x \cdot (q^2 - q^1) \times (q^2 - q^1) \geq 0. \end{aligned}$$

This optimization problem is similar to the optimization problem that is formulated in the article, except that the equality constraints are penalized in the article and the non-penetrating constraint is here left out. The objective function is minimized if the normals point in parallel but in opposite directions which corresponds to a complete fold. The equality constraints are used to maintain the correct shape of the sheets, and the inequality constraint, which is the triple product, corresponds to the signed volume of the tetrahedron spanned by the four vertices. The inequality constraint forces the fold to go in one direction, instead of two. In the article the inequality constraint is used to avoid "twists".

If we insert the data, the problem transforms into

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & -x_2, \\ \text{subject to} \quad & \|x\|^2 = 1, \\ & \|x - (1,0,0)\|^2 = 2, \\ & x_3 \geq 0. \end{aligned}$$

The problem is obviously not convex, but the constraint qualification MFCQ holds, and thus the global optimal solution (there are no other local optima) $x^* = (0,1,0)$ is a KKT point.

The reason that the equality constraints above are penalized in the formulation in the article is that, with a general triangulation of the faces, there is a risk that the edges give rise to linear dependent constraints (failure of LICQ and algorithms).

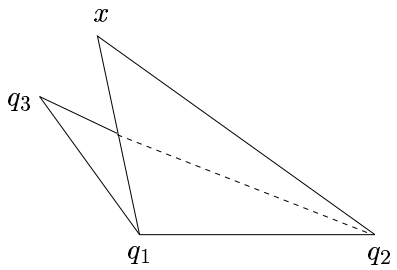


Figure 7: Fold of two triangular faces.

3.4.3 Future work

A few months after the Licentiate report was written, we developed an alternative numerical algorithm for folding airbags. The algorithm is more direct and faster than the one presented in the paper. The end result is also not identical, which in fact may be an advantage. The folded state from the previous method consisted of several non-planar connected patches. Once the folded state is found, usually other algorithms are used to fold or roll the airbag yet again to fit it into its compartment. These algorithms may fail depending on the orientation of the patches. The new algorithm generates patches that either lie in the xy-plane or are orthogonal to it. All patches will be perfectly flat. The downside is that the area of the full airbag is not equal to the flattened airbag.

The method is organized in a series of steps. The polyhedron is cut (and separated) along each crease. Each patch is rotated and placed at a specific height that has previously been determined. In the final step, the polyhedron is connected with additional patches, which gives a reasonable approximation if the thickness of the airbag is small.

The crease pattern splits the faces of the polyhedron into smaller patches. We let each patch be represented by a node in a directed graph. We put an arc between two nodes in the graph if the two patches on the polyhedron share a crease. The arc is directed to the node (patch) that will be above the other in the folded flat state of the polyhedron. To each arc $j \in J := \{1, \ldots, n\}$ in the graph, we assign a weight $z_j \in \mathbb{R}_+$. The weights will determine the height of the folded polyhedron in each position.

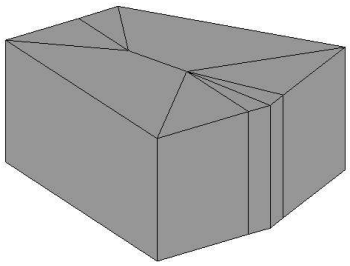


Figure 8: A polyhedron with crease pattern. The crease subdivide the faces into patches.

Each vertex (intersection of creases) in the polyhedron results in a cycle in the graph. We assign an equation to each vertex: the sum of the weights in the cycle, traveled in one direction, should be zero (traveling along an arc which points in the opposite direction yields a negative weight). Each weight is constrained to be larger than or equal to the artificial thickness τ of the polyhedron.

We assume that there are m vertices (intersection of creases) in the polyhedron and we let $I := \{1, \ldots, m\}$. We formulate the folding problem as a linear program, where we wish to minimize the total height:

$$\begin{aligned} \underset{z}{\text{minimize}} \quad & \sum_{j \in J} z_j \\ \text{subject to} \quad & \sum_{j \in J_1^i} z_j - \sum_{j \in J_2^i} z_j = 0, \qquad i \in I \\ & z_j \geq \tau, \qquad j \in J, \end{aligned}$$

where J_1^i and J_2^i denote the arcs that point in positive and negative direction corresponding to vertex i, respectively. The optimal solution z^* is used to place each patch at the correct height. The idea is that once a patch is fixed to a certain height, adjacent patches are positioned (in height) according to the weight of the edge joining the two patches.

Once the height of each patch is determined, we rotate each patch in sequence. First we consider the patch at the bottom of the polyhedron and place it in the xy-plane with $z = 0$. Next we put all adjacent patches in a stack for subsequent rotation. In each iteration we pop the stack for a patch (called parent) and rotate it. We also place all its adjacent patches (children) onto the stack (if not previously imported onto the stack). The x and y coordinates for patch j is determined by the orientation of the parent

and the crease. The z coordinate is the sum of the z coordinate for the parent and the optimal solution z_j^*. Figure 9 shows the rotated patches of the polyhedron in Figure 8.

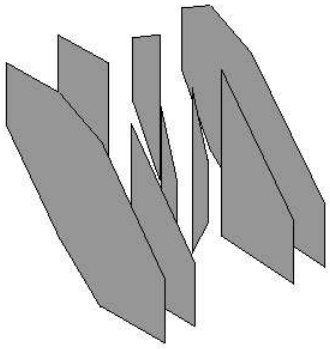


Figure 9: Rotated patches of the polyhedron in Figure 8. The "bottom" patch is the closest patch. The size of the separation of the patches (artificial thickness) is exaggerated.

At this stage, all patches will lie parallel to the xy-plane but with different z coordinates. We then connect the patches with additional patches that will be orthogonal to the rotated patches. Figure 10 shows the connections for the fold in Figure 9. Figure 11 shows a polyhedron with a crease pattern

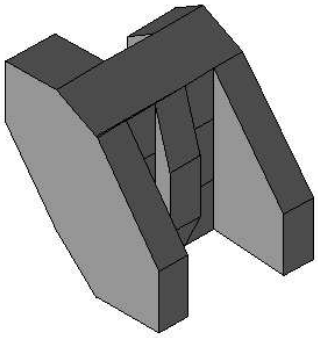


Figure 10: Connections (dark color) for the folded polyhedron from Figure 9. The size of the separation is exaggerated.

that has been generated according to Paper IV. The folded model is shown in Figure 12. The algorithm and a graphical user interface (see Figure 13) have been implemented in MATLAB.

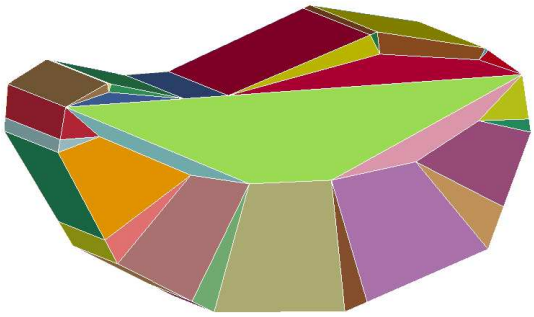


Figure 11: Model of a passenger airbag with a crease pattern.

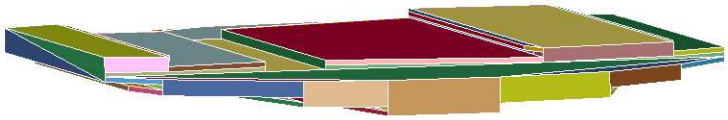


Figure 12: Folded model of passenger airbag. The thickness is exaggerated.

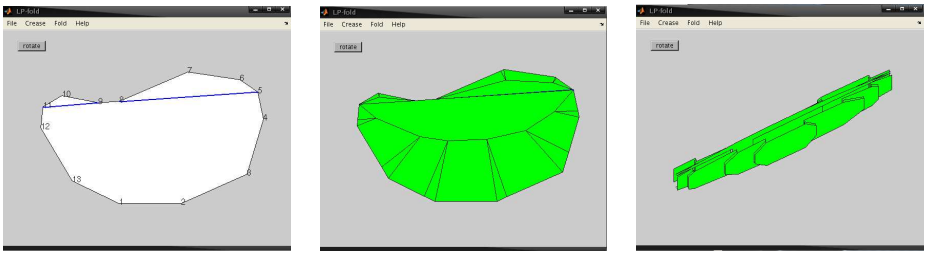
References

[Ber82] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Computer Science and Applied Mathematics. Academic Press Inc., New York, 1982.

[BL97] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming.* Springer Series in Operations Research. Springer-Verlag, New York, 1997.

[Bor06] T. Bortfeld. IMRT: a review and preview. *Phys. Med. Biol.*, 51:363–379, 2006.

[Bra95] A. Brahme. Treatment optimization using physical and radiobiological objective functions. In A. R. Smith, editor, *Radiation Therapy Physics*, pages 209–246, Berlin, 1995. Springer–Verlag.



(a) (b) (c)

Figure 13: Graphical user interface for importing/exporting geometries and folding.

[BRB08] B. T. Baumrucker, J. G. Renfro, and L. T. Biegler. MPEC problem formulations and solution strategies with chemical engineering applications. *Comp. & Chem. Eng.*, 32(12):2903–2913, 2008.

[BSS06] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2006.

[BT89] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[BT95] P. T. Boggs and J. W. Tolle. Sequential quadratic programming. In *Acta Numerica 1995*, pages 1–51. Cambridge Univ. Press, Cambridge, 1995.

[BTN02] A. Ben-Tal and A. Nemirovski. Robust optimization: methodology and applications. *Math. Program.*, 92(3, Ser. B):453–480, 2002.

[Car08] J. Carlsson. *Utilizing Problem Structure in Optimization of Radiation Therapy*. PhD thesis, Department of Optimization and Systems Theory, Royal Institute of Technology, Stockholm, 2008.

[CBT06] T. C. Y. Chan, T. Bortfeld, and J. N. Tsitsiklis. A robust approach to IMRT optimization. *Phys. Med. Biol.*, 51:2567–2583, 2006.

[CD02] B. Choi and J. O. Deasy. The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning. *Phys. Med. Biol.*, 47:3579–3589, 2002.

[CZHS05] M. Chu, Y. Zinchenko, S. G. Henderson, and M. B. Sharpe. Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. *Phys. Med. Biol.*, 50:5463–5477, 2005.

[DAPM00] K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In *Proceedings of the Parallel Problem Solving from Nature VI Conference*, pages 849–858, Paris, France, 2000. Springer-Verlag.

[DBC03] J. Deasy, A. I. Blanco, and V. H. Clark. CERR: A computational environment for radiotherapy research. *Med. Phys.*, 30(5):979–985, 2003.

[DD97] I. Das and J. E. Dennis Jr. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Structural Optimization*, 14:63–69, 1997.

[DR09] A. L. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings.* Springer-Verlag, New York, 2009.

[Ehr05] M. Ehrgott. *Multicriteria Optimization.* Springer-Verlag, Berlin, second edition, 2005.

[EP03] A. Evgrafov and M. Patriksson. Stable relaxations of stochastic stress-constrained weight minimization problems. *Struct. Multidiscip. Optim.*, 25(3):189–198, 2003.

[EPP03] A. Evgrafov, M. Patriksson, and J. Petersson. Stochastic structural topology optimization: existence of solutions and sensitivity analyses. *ZAMM Z. Angew. Math. Mech.*, 83(7):479–492, 2003.

[FGW02] A. Forsgren, P. E. Gill, and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Rev.*, 44(4):525–597, 2002.

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

[FLRS06] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes. Local convergence of SQP methods for mathematical programs with equilibrium constraints. *SIAM J. Optim.*, 17(1):259–286, 2006.

[FLT02] R. Fletcher, S. Leyffer, and Ph. L. Toint. On the global convergence of a filter-SQP algorithm. *SIAM J. Optim.*, 13(1):44–59, 2002.

[FP03a] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems. Vol. I.* Springer Series in Operations Research. Springer-Verlag, New York, 2003.

[FP03b] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems. Vol. II.* Springer Series in Operations Research. Springer-Verlag, New York, 2003.

[Geo68] A. M. Geoffrion. Proper efficiency and the theory of vector maximization. *J. Math. Anal. Appl.*, 22:618–630, 1968.

[GMS05] P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: an SQP algorithm for large-scale constrained optimization. *SIAM Rev.*, 47(1):99–131, 2005.

[GOT04] N. I. M. Gould, D. Orban, and Ph. L. Toint. GALAHAD, a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization. *ACM Trans. Math. Software*, 29(4):353–372, 2004.

[HK92] B. F. Hobbs and K. A. Kelly. Using game theory to analyze electric transmission pricing policies in the united states. *European J. Oper. Res.*, 56:154–171, 1992.

[KTY+06] K. Kuribayashi, K. Tsuchiya, Z. You, D. Tomus, and M. Umemoto. Self-deployable origami stent grafts as a biomedical application of Ni-rich TiNi shape memory alloy foil. *Materials Sci. Eng.:A*, 419:131–137, 2006.

[LÖ0] J. Löf. *Development of a General Framework for Optimization of Radiation Therapy.* PhD thesis, Department of Medical Radiation Physics, Stockholm University, 2000.

[Lan03] R. J. Lang. *Origami Design Secrets: Mathematical Methods for an Ancient Art.* AK Peters, Wellesley, MA, 2003.

[LLCN06] S. Leyffer, G. López-Calva, and J. Nocedal. Interior methods for mathematical programs with complementarity constraints. *SIAM J. Optim.*, 17(1):52–77, 2006.

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

[Mie99] K. Miettinen. *Nonlinear Multiobjective Optimization.* Kluwer's International Series in Operations Research & Management Science. Kluwer Academic Publishers, Boston, MA, 1999.

[MP07] P. Marcotte and M. Patriksson. Traffic equilibrium. In C. Barnhart and G. Laporte, editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, pages 623–713, Amsterdam, 2007. North–Holland.

[Mye69] J. H. Myer. Optigami: A tool for optical system design. *Applied Optics*, 8(2):260, 1969.

[NW06] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.

[OJW05] A. Olafsson, R. Jeraj, and S. J. Wright. Optimization of intensity-modulated radiation therapy with biological objectives. *Phys. Med. Biol.*, 50:5357–5379, 2005.

[OW06] A. Olafsson and S. J. Wright. Efficient schemes for robust IMRT treatment planning. *Phys. Med. Biol.*, 51:5621–5642, 2006.

[Pat94] M. Patriksson. *The Traffic Assignment Problem: Models and Methods.* VSP, Zeist, 1994.

[Pat08a] M. Patriksson. On the applicability and solution of bilevel optimization models in transportation science: A study on the existence, stability and computation of optimal solutions to stochastic mathematical programs with equilibrium constraints. *Transportation Research*, 42B(10):843–860, 2008.

[Pat08b] M. Patriksson. Robust bi-level optimization models in transportation science. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 366(1872):1989–2004, 2008.

[PW99] M. Patriksson and L. Wynter. Stochastic mathematical programs with equilibrium constraints. *Oper. Res. Lett.*, 25(4):159–167, 1999.

[Ral08] D. Ralph. Mathematical programs with complementarity constraints in traffic and telecommunications networks. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 366(1872):1973–1987, 2008.

[SG04] O. Schenk and K. Gärtner. Solving unsymmetric sparse systems of linear equations with PARDISO. *Journal of Future Generation Computer Systems*, 20(3):475–487, 2004.

[SG06] O. Schenk and K. Gärtner. On fast factorization pivoting methods for symmetric indefinite systems. *Elec. Trans. Numer. Anal.*, 23:158–179, 2006.

[She85] Y. Sheffi. *Urban Transportation Networks.* Prentice-Hall, Englewood Cliffs, 1985.

[SS00] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. *Math. Oper. Res.*, 25(1):1–22, 2000.

[VS52] H. Van Stackelberg. *The Theory of Market Economy.* Oxford University Press, Oxford, 1952.

[War52] J. G. Wardrop. Some theoretical aspects of road traffic research. In *Proceedings of the Institute of Civil Engineers, Part II*, pages 325–378, 1952.

[Web01] S. Webb. *Intensity-Modulated Radiation Therapy.* Institute of Physics Publishing, London, UK, 2001.

[WMNSU02] Q. Wu, R. Mohan, A. Niemierko, and R. Schmidt-Ullrich. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. *I. J. Radiation. Oncology. Biol. Phys.*, 52(1):224–235, 2002.

Paper I

On the Robustness of Global Optima and Stationary Solutions to Stochastic Mathematical Programs with Equilibrium Constraints, Part 1: Theory

C. Cromvik* M. Patriksson†

Communicated by F. Giannessi

Abstract We consider a stochastic mathematical program with equilibrium constraints (SMPEC), and show that, under certain assumptions, global optima and stationary solutions are robust with respect to changes in the underlying probability distribution. In particular, the discretization scheme sample average approximation (SAA), which is convergent for both global optima and stationary solutions, can be combined with the robustness results to motivate the use of SMPECs in practice. We then study two new and natural extensions of the SMPEC model. First, we establish the robustness of global optima and stationary solutions to an SMPEC model where the upper-level objective is the risk measure known as conditional value-at-risk (CVaR). Second, we analyze a multiobjective SMPEC model, establishing the robustness of weakly Pareto optimal and weakly Pareto stationary solutions. In the accompanying paper (Cromvik and Patriksson, On the Robustness of Global Optima and Stationary Solutions to Stochastic Mathematical Programs with Equilibrium Constraints, part 2: Applications, Journal of Optimization Theory and Applications, 2010, to appear) we present applications of these results to robust traffic network design and to robust intensity modulated radiation therapy.

*Department of Mathematical Sciences, Chalmers University of Technology and Department of Mathematical Sciences, University of Gothenburg, SE-412 96 Gothenburg, Sweden. E-mail: christoffer.cromvik@chalmers.se

†Corresponding author. Department of Mathematical Sciences, Chalmers University of Technology and Department of Mathematical Sciences, University of Gothenburg, SE-412 96 Gothenburg, Sweden. E-mail: mipat@chalmers.se

Keywords: Stochastic mathematical program with equilibrium constraints, Solution stability and robustness, Sample average approximation, Weak Pareto optimality

1 Introduction

A physical system is often subjected to uncertainties, such as uncertain material parameters in a structural problem and stochastic market in an application to economics. Although it may be sufficient to consider a representative (e.g. mean) value of the uncertain parameters in order to produce a simulation of the system, the solution to an optimization problem that is based on its response can be very sensitive; using mean values of uncertain parameters can then give suboptimal solutions.

There are two main approaches to incorporating uncertainty into optimization models. In stochastic programming (see e.g. [1, 2]), the objective function is an expected value, sometimes in combination with a risk measure. In such models, either all the decisions are made before any realization of the uncertain data, or the model contains a recourse opportunity so that some decisions can be made at a second stage. In robust optimization, an optimal solution is required to be feasible for all realizations of the uncertain data. The probability distributions utilized in stochastic programming are here replaced by the requirements that parameter values are confined to special bounded sets. Robust optimization provides a guarantee that an optimal solution is safe, but it is a pessimistic approach, since it considers the worst-case scenario. It is tractable for certain convex problems (see [3]); however, it has also been utilized for nonconvex problems through a linearization of the constraints ([4]).

We are interested in several applications which all can be modeled as a mathematical program with equilibrium constraints (MPEC). The stochastic extension of MPEC is a stochastic mathematical program with equilibrium constraints (SMPEC) ([5]); the focus of this paper is to analyze the stability of optimal and stationary solutions to SMPEC when the underlying probability distribution is itself uncertain. This is motivated by practical applications such as the optimization of treatment plans in radiation therapy, where the biological response to radiation depends on parameters of which the probability distribution is uncertain; see our further discussions on this topic in [6].

We show that the SMPEC model is robust under certain conditions; we also show how to discretize a continuous distribution using sample average approximation (SAA) and that such an approximation will converge. Not

surprisingly, similar conditions are required for robustness as for convergence of SAA.

The quantitative stability of solutions to stochastic programs due to changes in the probability distribution has been studied previously in [7] for general stochastic programs, for convex programs in [8], and for multistage programs in [9]; their focus is on the Lipschitz continuity of global optimal objective values. In contrast, our approach is qualitative in nature and provides stability results on optimal as well as stationary solutions. The subjects of optimality conditions and numerical methods for SMPECs have been studied previously, for example, in [10–12].

The remainder of this paper is organized as follows. In Section 2, we introduce the SMPEC model. In Section 3, we derive conditions under which global optima and stationary solutions are stable with respect to perturbations in the probability distribution. In Section 4, we present an extension of the SMPEC to include the risk measure CVaR, and establish the robustness of its global optima and stationary solutions. In Section 5, we present a discretization scheme, and show that it converges when the discretization is refined. Combining stability with a convergent discretization scheme provides a motivation for the use of SMPEC in practice. In Section 6, we extend the stability result to SMPECs with multiple objectives. Finally, in Section 7, we provide a summary and future research opportunities.

2 Stochastic Mathematical Programs with Equilibrium Constraints

2.1 Mathematical Programs with Equilibrium Constraints

Consider a mathematical program with an equilibrium constraint in the form of a variational inequality,

$$\begin{array}{lll} \text{(MPEC)} & \min\limits_{(x,y)} & f(x,y), \\ & \text{s.t.} & x \in X, \\ & & -F(x,y) \in N_C(y), \end{array}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $y \in \mathbb{R}^m$, $C \subseteq \mathbb{R}^m$ is a polyhedron, $F(x,\cdot) : C \to \mathbb{R}^m$ is smooth, and $N_C : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is the standard normal cone mapping

$$N_C(y) := \begin{cases} \left\{ z \in \mathbb{R}^m \,|\, z^{\mathrm{T}}(w-y) \le 0,\ w \in C \right\}, & \text{if } y \in C, \\ \emptyset, & \text{otherwise.} \end{cases}$$

The vector $x \in \mathbb{R}^n$ represents the design (or primary) variables and $y \in \mathbb{R}^m$ is the response (or secondary) variables. The nonempty, closed and convex

set $X \subseteq \mathbb{R}^n$ specifies the set of feasible designs. Note that there are no joint upper-level constraints in this setting, which is natural when considering the stability of optimal solutions (cf. [13–15]).

The variational inequality, $-F(x,y) \in N_C(y)$, can represent an equilibrium in a general form. For example, with $C = \mathbb{R}^m$, the variational inequality is equivalent to the system of equations

$$F(x,y) = 0^m,$$

and with $C = \mathbb{R}^m_+$, the variational inequality is equivalent to the complementarity constraint

$$0^m \leq y \perp F(x,y) \geq 0^m,$$

where $a \perp b$ means that $a^{\mathrm{T}} b = 0$. Since complementarity constraints are examples of equilibrium constraints, it indicates that MPECs are usually very nonlinear and irregular. In fact, MPECs lack standard constraint qualifications ([16]), which can highly influence the performance of nonlinear optimization algorithms for solving MPECs. For recent work on the numerical solution of MPECs, we refer to [17, 18].

If $F(x,y) = \nabla_y \phi(x,y)$ for a C^1 function $\phi(x,\cdot)$, then the variational inequality

$$\nabla \phi(x,y)^{\mathrm{T}} (\bar{y} - y) \geq 0, \qquad \forall \bar{y} \in C,$$

represents the optimality conditions for the parametric optimization problem

$$\min_{y \in C} \phi(x,y),$$

and the MPEC becomes what is traditionally known as a bilevel optimization problem ([19]).

The generality of the variational inequality suggests that a number of optimization problems can be put into the form of (MPEC) (see [20–22]). For example, the Stackelberg game [23], which is a leader-follower game and an extension of the Nash game [24], can be formulated as an MPEC. The accompanying paper [6] numerically analyzes applications to the design of traffic networks and optimal treatment plans in intensity-modulated radiation therapy (IMRT).

2.2 Stochastic Mathematical Program with Equilibrium Constraints

Next, we consider the stochastic extension of (MPEC). Let (Ω, Θ, P) be a complete probability space and consider the problem

$$\begin{array}{lll} (\mathrm{SMPEC}_\Omega) & \min\limits_{(x,y(\cdot))} & \mathrm{E}_\omega[f(x, y(\omega), \omega)] := \displaystyle\int_\Omega f(x, y(\omega), \omega)\, P(d\omega), \\ & \text{s.t.} & x \in X, \\ & & -F(x, y, \omega) \in N_C(y), \qquad P\text{-a.s.}, \end{array}$$

where $y : \Omega \to \mathbb{R}^m$ is a random element of the probability space (Ω, Θ, P). We also introduce $S : \mathbb{R}^n \times \Omega \rightrightarrows \mathbb{R}^m$, which defines the set of solutions to the lower-level parametric variational inequality problem,

$$S(x, \omega) := \{\, y \in \mathbb{R}^m \mid -F(x, y, \omega) \in N_C(y) \,\}.$$

In view of stochastic programming with recourse, SMPEC is considered as a here-and-now type of problem, where decision x should be taken before any realizations of uncertain data.

When the solution to the lower-level problem is nonunique for a fixed x and ω, the model should be interpreted as y being chosen such that the objective function is minimized given x and ω.

This formulation of SMPEC follows the original one of Patriksson and Wynter [5] and of Evgrafov and Patriksson [25]. Alternative formulations are found in [10, 11, 26].

2.3 Existence of Optimal Solutions

The following assumption will be in force throughout this paper:

Assumption A

(A1) The mapping $S(x, \cdot)$ is measurable for every x.

(A2) The set X is closed and the mapping $x \mapsto S(x, \omega)$ is closed for almost every $\omega \in \Omega$.

(A3) The function f is continuous in (x, y), measurable in ω, uniformly weakly coercive with respect to x over the set X, and bounded from below by a (Θ, P)-integrable function.

(A4) The set $S(x_0, \omega)$ is nonempty for some $x_0 \in X$ and almost every $\omega \in \Omega$.

The following result on the existence of optimal solutions is due to Evgrafov and Patriksson [25].

Theorem 2.1 (Existence of Optimal Solutions) *Let Assumption* (A) *hold. Then, problem* (SMPEC_Ω) *has at least one optimal solution.*

3 Solution Stability

We are interested in the stability of optimal solutions to (SMPEC_Ω) with respect to changes in the probability distribution. Such results have value both from a computational and a theoretical viewpoint. If the problem is stable, then it can be approximated by using discrete probability measures, resulting in a finite-dimensional problem. From a theoretical point of view, we deduct that the problem is robust. We analyze the stability of globally optimal solutions (Theorem 3.1) as well as of stationary solutions (Theorem 3.2).

3.1 Stability of Globally Optimal Solutions

We first analyze the stability of globally optimal solutions. This is foremost motivated by simplicity. For global optima, we can relate the perturbations of the probability distribution to changes in the objective value. For stationary solutions, this becomes more problematic.

The result of stability of globally optimal solutions is particularly interesting for convex problems, where we can find global optima in practice.

Let $\{P_k\}$ be a sequence of probability measures defined on $\mathcal{B}(\Omega)$. Consider the associated sequence of optimization problems,

$$
\begin{array}{lll}
(\text{SMPEC}_\Omega)^k & \min\limits_{(x,y(\cdot))} & \mathrm{E}_\omega[f(x,y(\omega),\omega)] := \displaystyle\int_\Omega f(x,y(\omega),\omega)\, P_k(d\omega), \\
& \text{s.t.} & x \in X, \\
& & -F(x,y,\omega) \in N_C(y), \qquad P_k\text{-a.s.},
\end{array}
$$

The problem differs from (SMPEC_Ω) only in the choice of probability distribution. Let $\text{val}(P)$ denote the optimal value of problem P. The following result shows the stability of globally optimal solutions. The corresponding result in the context of topology optimization in structural mechanics can be found in [15] and for network design under traffic equilibrium in [27, 28]. The proof presented here is similar.

Theorem 3.1 (Global Stability of Optimal Solutions to (SMPEC_Ω)) *Let Assumption* (A) *hold, suppose that the mapping $F(x,\cdot,\omega)$ is strictly monotone in y for each $x \in X$ and $\omega \in \Omega$, and that the sequence $\{P_k\}$ of probability measures weakly converges to P. Also, suppose that, for each k, $(x^k, y^k(\cdot))$ solves ($\text{SMPEC}_\Omega)^k$. Then, each limit point (there is at least one) of the sequence $\{(x^k, y^k(\cdot))\}$ is an optimal solution to* (SMPEC_Ω).

Proof. Consider an optimal solution $(x^*, y^*(\cdot))$ to (SMPEC_Ω). Since F is strictly monotone in y, y is continuous in x and ω. By (A3), any sequence of feasible designs and responses is bounded and hence has a limit point. The optimal solution to (SMPEC_Ω) is moreover feasible in $(\text{SMPEC}_\Omega)^k$ for all k. It follows that $\text{val}\,(\text{SMPEC}_\Omega) \geq \limsup_{k\to\infty} \text{val}\,(\text{SMPEC}_\Omega)^k$.

Next, let $\{(x^k, y^k(\cdot))\}$ be a sequence of optimal solutions to $(\text{SMPEC}_\Omega)^k$. By (A3), this sequence is bounded. Denote any limit point $(\bar{x}, \bar{y}(\cdot))$. It is feasible for almost every ω in (SMPEC_Ω). Using the lower semicontinuity of f and Fatou's lemma, we get

$$\begin{aligned}
\text{val}\,(\text{SMPEC}_\Omega) &\leq \int_\Omega f(\bar{x}, \bar{y}(\omega), \omega) p(\omega)\, d\omega \\
&\leq \int_\Omega \liminf_{k\to\infty} f(x^k, y^k(\omega), \omega) p(\omega)\, d\omega \\
&\leq \liminf_{k\to\infty} \int_\Omega f(x^k, y^k(\omega), \omega) p_k(\omega)\, d\omega \\
&= \liminf_{k\to\infty} \text{val}\,(\text{SMPEC}_\Omega)^k.
\end{aligned}$$

By combining the two inequalities, we get the result. □

3.2 Stability of Stationary Solutions

Due to the nonconvex nature of MPECs, it is not reasonable in general to expect algorithms to find globally optimal solutions. This fact limits the practical use of Theorem 3.1, and raises the question of stability of stationary solutions. The proof of stability for globally optimal solutions was based on analyzing the convergence of the optimal value. For stationary solutions, we need to analyze the conditions of (local) optimality which relates to stationarity.

Optimality conditions for (SMPEC_Ω) are nontrivial to formulate due to the presence of the variational inequality. Under certain conditions, the response variable y can be treated as an implicit variable; this reduces the complexity of formulating optimality conditions for the SMPEC. This technique is used by Outrata [21] for the MPEC, which has inspired the proof approach below.

The following assumption will be utilized in addition to Assumption A.

Assumption B

(B1) The function f is Lipschitz continuous in (x, y).

(B2) The mapping $F(\cdot, \cdot, \omega)$ is continuously differentiable and $F(x, \cdot, \omega)$ is uniformly strongly monotone on C for each $x \in X$ and $\omega \in \Omega$, i.e.,

$$(F(x, y_1, \omega) - F(x, y_2, \omega))^{\mathrm{T}}(y_1 - y_2) \geq c\|y_1 - y_2\|^2, \quad \forall y_1, y_2 \in C,$$

where $c > 0$ is independent of x and ω.

(B3) $X = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0,\ i = 1, ..., p\}$ and each function g_i is continuously differentiable;

(B4) The Mangasarian–Fromovitz constraint qualification (MFCQ) holds for all $x \in X$.

If Assumptions (B1) and (B2) hold, then (see [29]) there exists a locally Lipschitz continuous, single-valued solution map $(x, \omega) \mapsto \sigma(x, \omega)$ with

$$y = \sigma(x, \omega), \qquad \sigma(x, \omega) \in S(x, \omega).$$

With this property, we can rewrite (SMPEC_Ω) as the one-level problem,

$$(\mathrm{SNLP}_\Omega) \quad \min_x \ \mathrm{E}_\omega[f(x, \sigma(x, \omega), \omega)] := \int_\Omega f(x, \sigma(x, \omega), \omega)\, P(d\omega),$$
$$\text{s.t.} \quad x \in X,$$

and correspondingly $(\mathrm{SNLP}_\Omega)^k$ is obtained from (SMPEC_Ω) by replacing P with P_k.

Before stating the optimality conditions, we introduce two definitions from nonsmooth analysis.

Definition 3.1 *The Clarke directional derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ at x in the direction h is defined by*

$$f^0(x; h) := \limsup_{\substack{t \downarrow 0 \\ z \to x}} \frac{f(z + th) - f(z)}{t}.$$

Since f is Lipschitz continuous [Assumption (B1)], $f^0(\cdot, h)$ is upper semicontinuous ([30, Proposition 2.1.1]).

Definition 3.2 *The generalized gradient of f at x is defined as the set*

$$\partial f(x) := \{\xi \in \mathbb{R}^n \mid (\xi, h) \leq f^0(x; h)\}.$$

If f is continuously differentiable at x then $\partial f(x) = \{\nabla f(x)\}$. If Assumptions (B3) and (B4) hold, a vector $x^* \in X$ is a Clarke stationary solution (see [30, Theorem 6.1.1], see [31, Theorem 6.1.8]) to (SNLP_Ω) if, for some vector $\mu \in \mathbb{R}^p_+$ with $\mu_i g_i(x^*) = 0$ for all i, we have

$$0^n \in \partial \mathrm{E}_\omega[f(x^*, \sigma(x^*, \omega), \omega)]) + \nabla g(x^*)\mu,$$

where μ is the vector of Lagrange multipliers.

Theorem 3.2 (Stability of Stationary Solutions to (SNLP_Ω)) *Let Assumption* (A) *and* (B) *hold, suppose that the sequence $\{P_k\}$ of probability measures weakly converges to P and is upper bounded by a measurable function, and that, for each k, $(x^k, y^k(\cdot))$ is a Clarke stationary solution to* $(\mathrm{SNLP}_\Omega)^k$. *Then, each limit point (there is at least one) of the sequence $\{(x^k, y^k(\cdot))\}$ is a Clarke stationary solution to* (SNLP_Ω).

Proof. By the assumptions, $F(x, \cdot, \omega)$ is uniformly strongly monotone, and therefore the solution map $S(x, \omega)$ is single-valued and Lipschitz continuous (cf. [29]). This enables us to use the one-level problems (SNLP_Ω) and $(\mathrm{SNLP}_\Omega)^k$. Since σ and f are Lipschitz continuous, there exists a random variable $k(\omega) \geq 0$ such that $\mathrm{E}[k(\omega)] < \infty$ and such that, for all $x_1, x_2 \in X$,

$$|f(x_1, \sigma(x_1, \omega), \omega) - f(x_2, \sigma(x_2, \omega), \omega)| \leq k(\omega)\|x_1 - x_2\|. \tag{1}$$

Let x^k be a Clarke stationary solution to $(\mathrm{SNLP}_\Omega)^k$ and consider a sequence $\{x^k\}$ of such stationary solutions. Since f is inf-compact [see Assumption (A3)], this sequence is bounded. Denote a limit point by x^*. Define

$$\mathrm{E}^k_\omega[f](x) := \int_\Omega f(x, \sigma(x, \omega), \omega)\, P_k(d\omega).$$

The point x^k is stationary if

$$0^n \in \partial \mathrm{E}^k_\omega[f](x^k) + \nabla g(x^k)\mu^k \tag{2}$$

and

$$0^p \leq \mu^k \perp g(x^k) \leq 0^p.$$

Fix a direction $h \in \mathbb{R}^n$. Then, we have

$$\begin{aligned}
(\mathrm{E}_\omega^k[f])^0(x;h) &= \limsup_{\substack{t\downarrow 0\\ z\to x}} \frac{\mathrm{E}_\omega^k[f](z+th) - \mathrm{E}_\omega^k[f](z)}{t} \\
&= \limsup_{\substack{t\downarrow 0\\ z\to x}} \left(\frac{\int_\Omega f(z+th, \sigma(z+th,\omega),\omega)\, P_k(d\omega)}{t} \right. \\
&\qquad \left. - \frac{\int_\Omega f(z,\sigma(z,\omega),\omega)\, P_k(d\omega)}{t} \right) \\
&= \limsup_{\substack{t\downarrow 0\\ z\to x}} \int_\Omega \frac{(f(z+th,\sigma(z+th,\omega),\omega) - f(z,\sigma(z,\omega),\omega))}{t}\, P_k(d\omega) \\
&\le \int_\Omega f^0(x,\sigma(x,\omega),\omega;h)\, P_k(d\omega) = E_\omega^k\left[f^0(x,\sigma(x,\omega),\omega;h)\right],
\end{aligned}$$

where the last inequality follows by Equation (1), with $x_1 = z$ and $x_2 = z + th$, and the Lebesgue dominated convergence theorem. Furthermore, we have that

$$\begin{aligned}
\limsup_{k\to\infty} (E_\omega^k[f])^0(x^k;h) &= \limsup_{k\to\infty} \int_\Omega f^0(x^k,\sigma(x^k,\omega),\omega;h)\, P_k(d\omega) \\
&\le \int_\Omega \limsup_{k\to\infty} f^0(x^k,\sigma(x^k,\omega),\omega;h)\, P_k(d\omega) \\
&\le \int_\Omega f^0(x^*,\sigma(x^*,\omega),\omega;h)\, P(d\omega) \\
&= E_\omega^0[f](x^*;h), \qquad (3)
\end{aligned}$$

where we use the Lebesgue dominated convergence theorem in the second equality and the upper semicontinuity of f^0 in the second inequality. Hence,

$$\limsup_{k\to\infty} \partial \mathrm{E}_\omega^k[f](x^k) \subset \partial \mathrm{E}_\omega[f](x^*).$$

Next, we argue that

$$\limsup_{k\to\infty} \nabla g(x^k)\mu^k = \nabla g(x^*)\mu^*.$$

Suppose that this is not true, but $\|\mu^k\| \to \infty$. We can then define $\lambda^k := \mu^k/\|\mu^k\|$, and assume that $\lambda^k \to \lambda^*$ for some λ^* such that

$$\lambda^* \ge 0^p,\ \|\lambda^*\| = 1.$$

Let $I(x) := \{i \mid g_i(x) = 0\}$ be the set of active constraint indices at x. By the definition of MFCQ, to each x^k, there exists a vector $d \in \mathbb{R}^n$ such that

$$\begin{aligned}
&\nabla g_i(x^k)^{\mathrm{T}} d < 0, \quad && i \in I(x^k), \\
&g_i(x^k) < 0, && i \notin I(x^k), \qquad (4)
\end{aligned}$$

For $i \in I(x^*)$ we must have $\lambda_i^k \to 0$. Fix a $d \in \mathbb{R}^n$ such that a condition like (4) holds at x^*. Then, by (2), we have

$$0 \leq \limsup_{k\to\infty} \left(\frac{d^{\mathrm{T}} \nabla g(x^k)\mu_k}{\|\mu_k\|} + \frac{(E_\omega^k[f])^0(x^k;d)}{\|\mu_k\|} \right) = \limsup_{k\to\infty} d^{\mathrm{T}} \nabla g(x^k)\lambda^k$$
$$= \sum_{i\in I(x^*)} \lambda_i^* \nabla g_i(x^*)^{\mathrm{T}} d.$$

From the last expression and (4), we get

$$0 \leq \sum_{i\in I(x^*)} \lambda_i^* \nabla g_i(x^*)^{\mathrm{T}} d \leq - \sum_{i\in I(x^*)} \lambda_i^*,$$

and since $\lambda^* \geq 0^p$, this implies that $\lambda^* = 0$. This contradicts the assumption $\|\lambda^*\| = 1$, and so μ^* must be bounded.

To sum up, we have

$$0 \in \limsup_{k\to\infty} \left(\partial \mathrm{E}_\omega^k[f](x^k) + \nabla g(x^k)\mu^k \right) \subset \partial \mathrm{E}_\omega[f](x^*) + \nabla g(x^*)\mu^*$$

and we can therefore conclude that x^* is a Clarke stationary solution to (SNLP_Ω). □

4 Risk Objective Function

In this section, we assume that the objective function $f(x, y(\omega), \omega)$ measures a loss. The value-at-risk (VaR) at probability level β is denoted by β-VaR(x); it is the value for which the probability that f exceeds this value is β, i.e.,

$$\beta\text{-VaR}(x) = \min\{\, \gamma \mid P(f(x, y(\omega), \omega) \geq \gamma) \geq \beta \,\}.$$

Furthermore, the conditional value-at-risk, denoted β-CVaR(x), is the conditional expectation of loss, given that the loss is greater than β-VaR, i.e.,

$$\beta\text{-CVaR}(x) = \frac{1}{1-\beta} \int_{f(x,y(\omega),\omega)\geq \beta\text{-Var}} f(x, y(\omega), \omega)\, P(d\omega).$$

The parameter β determines the level of risk. If $\beta = 1$, CVaR equals the expected value; if $\beta = 0$, CVaR equals the maximal value of f; the CVaR formalism therefore introduces the possibility to include interesting compromises between these two extremes.

Rockafellar and Uryasev [32] provide an alternative expression for β-CVaR which utilizes the following function:

$$G_\beta(x, y, \gamma) = \gamma + \frac{1}{1-\beta} \int_\Omega [f(x, y(\omega), \omega) - \gamma]_+ \, P(d\omega),$$

where, for $s \in \mathbb{R}$, $[s]_+ := \max\{0, s\}$. The conditional value-at-risk is equal to the minimum value of G_β over $\gamma \in \mathbb{R}$, i.e.,

$$\beta\text{-CVaR}(x) = \min_{\gamma \in \mathbb{R}} G_\beta(x, y, \gamma).$$

The value-at-risk is a minimizer of G_β, and the problem of minimizing β-CVaR over (x, y) is equivalent to minimizing G_β over (x, y, γ) ([32]). Note also that G_β is convex in γ, so CVaR preserves convexity.

Consider now the SMPEC where the expected value in the objective function is replaced by the alternative expression for β-CVaR:

$$\begin{array}{lll} (\text{SRPEC}_\Omega) & \min\limits_{(x, y(\cdot), \gamma)} & \gamma + \dfrac{1}{1-\beta} \displaystyle\int_\Omega [f(x, y(\omega), \omega) - \gamma]_+ \, P(d\omega), \\ & \text{s.t.} & x \in X, \\ & & -F(x, y, \omega) \in N_C(y), \qquad P\text{-a.s.}, \end{array}$$

We will next show that the robustness results for global optima (Theorem 3.1) and for stationary solutions (Theorem 3.2) can be extended to the SRPEC. Following the procedure in Section 3, let $\{P_k\}$ be a sequence of probability measures defined on $\mathcal{B}(\Omega)$ and consider the associated sequence of optimization problems $(\text{SRPEC}_\Omega)^k$ which are obtained from (SRPEC_Ω) by replacing P with P_k.

Theorem 4.1 (Global Stability of Optimal Solutions to (SRPEC_Ω)) *Let Assumption* (A) *hold, suppose that the mapping $F(x, \cdot, \omega)$ is strictly monotone in y for each $x \in X$ and $\omega \in \Omega$, and that the sequence $\{P_k\}$ of probability measures weakly converges to P. Also suppose that, for each k, $(x^k, y^k(\cdot), \gamma^k)$ solves $(\text{SRPEC}_\Omega)^k$. Then, each limit point (there is at least one) of the sequence $\{(x^k, y^k(\cdot), \gamma)\}$ is an optimal solution to* (SRPEC_Ω).

Proof. The proof follows essentially from that of Theorem 3.1. Two critical steps need to be motivated. The first is the continuity of the objective function. The function in the integral is continuous with respect to x and y, since it is a decomposition of f and $[\cdot]_+$, and since F is strictly monotone. Also, the objective function G_β is continuous with respect to γ.

The second step is that the objective function is weakly coercive with respect to (x, y, γ). This holds by Assumption (A3) and the fact that, if $|\gamma| \to \infty$, $G_\beta \to \infty$. □

To establish robustness for stationary solutions, we again utilize the reformulation of SMPEC into one-level problems. We consider the following problem:

$$\begin{array}{lll} (\text{SRNLP}_\Omega) & \min\limits_{(x, \gamma)} & \gamma + \dfrac{1}{1-\beta} \displaystyle\int_\Omega [f(x, \sigma(x, \omega), \omega) - \gamma]_+ \, P(d\omega), \\ & \text{s.t.} & x \in X. \end{array}$$

Correspondingly, $(\text{SRNLP}_\Omega)^k$ is obtained from (SRNLP_Ω) by replacing P with P_k.

Theorem 4.2 (Stability of Stationary Solutions to (SRNLP_Ω)) *Let Assumption* (A) *and* (B) *hold, suppose that the sequence $\{P_k\}$ of probability measures weakly converges to P and is upper bounded by a measurable function and that, for each k, $(x^k, y^k(\cdot), \gamma^k)$ is a Clarke stationary solution to $(\text{SRNLP}_\Omega)^k$. Then, each limit point (there is at least one) of the sequence $\{(x^k, y^k(\cdot), \gamma)\}$ is a Clarke stationary solution to (SRNLP_Ω).*

Proof. The proof follows from that of Theorem 3.2, the proof of Theorem 4.1, and the fact that G_β is Lipschitz continuous. □

5 Convergence of a Discretization Scheme

In this section, we discuss the numerical solution of (SMPEC_Ω). The objective function is a multidimensional integral which must be approximately computed in the general case. If it is discretized, it is natural to analyze the convergence of a discretization scheme. Having established both the stability of optimal solutions and the convergence of a numerical scheme puts us closer to the practical use of an SMPEC model.

One approach used extensively to numerically solve stochastic programs is a Monte Carlo technique known as sample average approximation (SAA) (see e.g. [33]). The idea is to draw N iid samples $\omega^1, \ldots, \omega^N$ and solve a deterministic problem for increasing values of N. We use the reformulation of $(\text{SMPEC})_\Omega$ into (SNLP_Ω) and consider the problem

$$(\text{SNLP})^N \quad \min_x \quad \hat{f}_N := \frac{1}{N}\sum_{k=1}^{N} f(x, \sigma(x, \omega^k), \omega^k),$$
$$\text{s.t.} \quad x \in X,$$

To establish convergence as $N \to \infty$, the following additional conditions are required.

Assumption C

(C1) The set X is bounded and convex.

(C2) The function $f(\cdot, \sigma(\cdot, \omega), \omega)$ is regular (i.e., f is directionally differentiable and the directional derivative coincides with the Clarke directional derivative) at x for almost every $\omega \in \Omega$.

Condition (C2) is fulfilled if e.g. $f(\cdot, \sigma(\cdot, \omega), \omega)$ is convex or continuously differentiable ([30]).

The convergence proofs for both globally optimal solutions and stationary solutions are based on the law of large numbers. The main difference in the assumptions needed is the requirement of a regular function for the convergence of stationary solutions.

Theorem 5.1 (Convergence of Optimal Solutions to (SNLP_Ω)) *Let Assumptions* (A), (B1)–(B2), (C1) *hold. For each N, let $(x_N, y_N(\cdot))$ be an optimal solution to* $(\mathrm{SNLP})^N$. *Then, each limit point (there is at least one) of the sequence $\{x_N\}$ is an optimal solution to* (SNLP_Ω).

Proof. The feasible set is compact by Assumptions (A2) and (C1); by Assumptions (B1)–(B2), for almost every $\omega \in \Omega$ the objective function $f(\cdot, \cdot, \omega)$ is continuous. By Assumption (B2), it is also bounded from above by a (Θ, P)-integrable function and the sample is iid. Then, by [33, Proposition 7], $\hat{f}_N$ converges to f w.p.1 uniformly on X. In turn, by [33, Proposition 5], this implies that $\mathrm{val}((\mathrm{SNLP})^N) \to \mathrm{val}((\mathrm{SNLP})_\Omega)$ as $N \to \infty$. □

Theorem 5.2 (Convergence of Stationary Solutions to (SNLP_Ω)) *Let Assumptions* (A), (B), (C) *hold. For each N, let $(x_N, y_N(\cdot))$ be a stationary solution to* $(\mathrm{SNLP})^N$. *Then, each limit point (there is at least one) of the sequence $\{x_N\}$ is a stationary solution to* (SNLP_Ω).

Proof. By Assumptions (A3) and (B2), the objective function is of Carathéodory type; by Assumption (C1), the set is compact and convex. By (C2), the function is also regular. Then, by [34, Theorem 7], the sequence of stationary solutions $\{x^N\}$ converges w.p.1 to a stationary solution of (SNLP_Ω). □

Since (SNLP_Ω) is a reformulation of (SMPEC_Ω), the above theorems state that we also have convergence to global optima and stationary solutions for the corresponding discretized problem

$$
\begin{aligned}
(\mathrm{SMPEC})^N \quad \min_x \quad & \hat{f}_N := \frac{1}{N} \sum_{k=1}^{N} f(x, y^k, \omega^k), \\
\text{s.t.} \quad & x \in X, \\
& -F(x, y^k, \omega^k) \in N_C(y^k), \qquad k = 1, \ldots, N.
\end{aligned}
$$

To summarize, the results from Theorems 5.1 and 5.2 show that it is a valid approach to compute a solution to the SMPEC model through a sequence of deterministic problems. Theorems 5.1 and 5.2 are also immediate to extend to the CVaR model of Section 4.

Analogous discretization schemes for engineering applications are studied in Evgrafov and Patriksson [14, 15].

6 Stochastic Multiple Objective Mathematical Programs with Equilibrium Constraints

With q objectives $f_i : X \to \mathbb{R}$, $i = 1, \ldots, q$, the standard multiobjective optimization problem is

$$\min_{x \in X} (f_1(x), \ldots, f_q(x)).$$

Let f denote the q-vector of functions f_i, $i = 1, \ldots, q$. We recall the definitions of Pareto and weakly Pareto optimal solutions.

Definition 6.1 *A vector $\bar{x} \in X$ is called* Pareto optimal *if there is no $x \in X$ such that $f(x) \leq f(\bar{x})$ and $f_i(x) < f_i(\bar{x})$ for at least one $i = 1, \ldots, q$. A feasible solution $\hat{x}$ is called* weakly Pareto optimal *if there is no $x \in X$ such that $f(x) < f(\hat{x})$.*

The study of a multiple objective SMPEC problem appears to be new; Ye and Zhu [35], and Murdukhovich [36,37] have studied multiobjective optimization versions of the MPEC problem. We define the multiple objective version of the SMPEC, the SMOPEC, for q objectives as

$$\begin{array}{lll} (\mathrm{SMOPEC}_\Omega) & \min\limits_{(x, y(\cdot))} & (\mathrm{E}_\omega[f_1(x, y(\omega), \omega)], \ldots, \mathrm{E}_\omega[f_q(x, y(\omega), \omega)]) , \\ & \text{s.t.} & x \in X, \\ & & -F(x, y, \omega) \in N_C(y), \qquad \text{P-a.s.} \end{array}$$

We analyze the stability of weakly Pareto optimal solutions (Theorem 6.1) and of weakly Pareto stationary solutions (Theorem 6.3) below.

6.1 Stability of Weakly Pareto-Optimal Solutions to a Convex Problem

Let $\{P_k\}$ be a sequence of probability measures defined on $\mathcal{B}(\Omega)$ and consider the associated sequence of optimization problems $(\mathrm{SMOPEC}_\Omega)^k$, which are obtained from (SMOPEC_Ω) by replacing P with P_k. If the set X is convex and the functions $f_i(\cdot, \cdot, \omega)$ are convex for $i = 1, \ldots, q$, then $(\mathrm{SMOPEC})_\Omega$ is a convex problem.

Theorem 6.1 (Stability of Weakly Pareto-Optimal Solutions to (SMOPEC_Ω)) *Let Assumption* (A) *hold, suppose that the mapping $F(x, \cdot, \omega)$ is strictly monotone for each $x \in X$ and $\omega \in \Omega$, that* (SMOPEC_Ω) *is a convex problem, and that the sequence $\{P_k\}$ of probability measures weakly converges to P. Also, suppose that for each k, $(x^k, y_k(\cdot))$ is a weakly Pareto-optimal solution to* $(\mathrm{SMOPEC}_\Omega)^k$. *Then, each limit point (there is at least one) of the sequence $\{(x^k, y^k(\cdot))\}$ is a weakly Pareto optimal solution to* (SMOPEC_Ω).

Proof. Consider a weakly Pareto optimal solution $(x^*, y^*(\cdot))$ to (SMOPEC_Ω). By convexity, there exists a vector $\lambda \in \mathbb{R}^q_+$ with $\lambda_i > 0$ for a least one $i = 1, \ldots, q$, such that the solution solves the following single-objective problem [38, Prop 3.10]:

$$\begin{array}{lll} \text{(S)} & \min\limits_{(x,y(\cdot))} & \sum\limits_{i=1}^{q} \int_\Omega \lambda_i f_i(x, y(\omega), \omega)\, P(d\omega), \\ & \text{s.t.} & x \in X, \\ & & -F(x, y, \omega) \in N_C(y), \qquad P\text{-a.s.} \end{array}$$

Fix the vector λ and consider a sequence of single-objective problems $(\text{S})^k$ which are obtained from (S) by replacing P with P_k.

Denote by $(x^k, y^k(\cdot))$ an optimal solution to $(\text{S})^k$. Since (SMOPEC_Ω) is a convex problem, so is $(\text{SMOPEC}_\Omega)^k$ for all k. By convexity, $(x^k, y^k(\cdot))$ is a weakly Pareto optimal solution to $(\text{SMOPEC}_\Omega)^k$ [38, Prop 3.9]. Now, apply Theorem 3.1 with the objective function f replaced by

$$f(x, y(\omega)) := \sum_{i=1}^{q} \lambda_i f_i(x, y(\omega)$$

to get the result. □

6.2 Stability of Weakly Pareto-Stationary Solutions

To establish stability without the assumption of convexity, we follow the development of Section 3.2 and reformulate (SMOPEC_Ω) and $(\text{SMOPEC}_\Omega)^k$ as one-level problems by treating y as a function of x and ω: $y = \sigma(x, \omega)$. This is possible if, in addition to the assumptions in Theorem 6.1, Assumptions (B1) and (B2) hold. We denote the reformulations by (SMONLP_Ω) and $(\text{SMONLP}_\Omega)^k$, respectively:

$$\begin{array}{lll} (\text{SMONLP}_\Omega) & \min\limits_{x} & (\text{E}_\omega[f_1(x, \sigma(x, \omega), \omega)], \ldots, \text{E}_\omega[f_q(x, \sigma(x, \omega), \omega)]), \\ & \text{s.t.} & x \in X, \end{array}$$

and

$$\begin{array}{lll} (\text{SMONLP}_\Omega)^k & \min\limits_{x} & (\text{E}^k_\omega[f_1(x, \sigma(x, \omega), \omega)], \ldots, \text{E}^k_\omega[f_q(x, \sigma(x, \omega), \omega)]), \\ & \text{s.t.} & x \in X, \end{array}$$

The following theorem is a KKT characterization of weak Pareto optimality for multiobjective problems due to Minami [39] and Li [40].

Theorem 6.2 *Consider the problem*

$$\begin{aligned} \min_x \quad & (f_1(x), \ldots, f_q(x)), \\ \text{s.t.} \quad & g_j(x) \leq 0, \quad j = 1, \ldots, p, \end{aligned}$$

where, for each $i = 1, \ldots, q$, f_i is locally Lipschitz continuous, and where for each $j = 1, \ldots, p$, $g_j \in C^1$. Let the MFCQ constraint qualification [Assumption (B4)] hold for all feasible solutions. Then, a feasible solution x^* is a weakly Pareto optimal solution if there exist real numbers $\lambda_i \geq 0$ for all i, with $\lambda_i > 0$ for at least one i, and a vector $\mu \in \mathbb{R}^p$ with $g_j(x^*)\mu_j = 0$ for $j = 1, \ldots, p$, such that

$$0^n \in \sum_{i=1}^{q} \lambda_i \partial f_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*).$$

A solution which fulfills these conditions is called a weakly Pareto-stationary solution.

Next, we establish the stability of weakly Pareto-stationary solutions.

Theorem 6.3
(Stability of Weakly Pareto-Stationary Solutions to (SMONLP$_\Omega$))
Let Assumptions (A) *and* (B) *hold, suppose that the sequence $\{P_k\}$ of probability measures weakly converges to P, and that, for each k, $(x^k, y^k(\cdot))$ is a weakly Pareto stationary solution to* (SMONLP$_\Omega$)^k. *Then, each limit point (there is at least one) of the sequence $\{(x^k, y^k(\cdot))\}$ is a weakly Pareto stationary solution to* (SMONLP$_\Omega$).

Proof. Consider a weakly Pareto-stationary solution x^* to (SMONLP$_\Omega$). By Theorem 6.2, there exist real numbers $\lambda_i \geq 0$ for all i, with $\lambda_i > 0$ for at least one i, such that the conditions in Theorem 6.2 hold. Fix this value of λ, and consider a sequence $\{(x^k, y^k(\cdot))\}$ which is Clarke stationary to (S)^k, that is,

$$0^n \in \partial \left(\sum_{i=1}^{q} \lambda_i f_i \right)(x^k) + \sum_{i=1}^{p} \mu_i \nabla g_i(x^k).$$

By the properties of the generalized gradient [30, Section 2.3, Corollary 2], we have that

$$\partial \left(\sum_{i=1}^{q} \lambda_i f_i \right)(x^k) \subset \sum_{i=1}^{q} \lambda_i \partial f_i(x^k),$$

for any scalars λ_i, $i = 1, \ldots, q$, so $(x^k, y^k(\cdot))$ is also a weakly Pareto stationary solution to $(\text{SMONLP}_\Omega)^k$. Now, apply Theorem 3.2 to the single-objective problem $(\text{S})^k$ with the objective function f replaced by

$$f(x, y(\omega)) := \sum_{i=1}^{q} \lambda_i f_i(x, y(\omega),$$

to get the result. □

7 Summary, Conclusions and Future Research

Our first and main contribution in this paper is that we established that the SMPEC model is robust under the assumptions that the solution to the lower-level equilibrium problem is unique and that we have sufficient regularity conditions on the objective function and constraints. We showed that global optima as well as stationary solutions are stable with respect to changes in the probability distribution. If the SMPEC framework is used to model the problem of finding a design which should be good on average for various scenarios, then the optimal solution to SMPEC gives a design which is stable to changing conditions. The result on the robustness also gives credibility to using stochastic programming in general, since one of the criticisms on stochastic programming is that the probability distribution is often unknown or only partially known.

Our second contribution is that we have formulated, and established the robustness of solutions to, two natural extensions of the SMPEC model: first, a model where the objective is the CVaR risk measure; second, a multiobjective SMPEC model.

We also presented a discretization scheme sample average approximation (SAA), which is convergent and can be used to solve the SMPEC model. The result on the convergence of the SAA scheme is not new, but was included to demonstrate that the results on robustness can be combined with a method for numerically solving the SMPEC model.

The accompanying paper [6] numerically analyzes applications of the SMPEC formalism to the design of traffic networks and optimal treatment plans in intensity-modulated radiation therapy (IMRT).

Acknowledgements This work was partially financially supported by the Gothenburg Mathematical Modeling Centre (GMMC) at Mathematical Sciences (Chalmers University of Technology and the University of Gothenburg), whose main sponsor is the Swedish Foundation for Strategic Research (SSF). We thank the Associate Editor Franco Giannessi for his very encouraging and constructive remarks on a previous manuscript.

References

[1] Kall, P., Wallace, S.W.: Stochastic programming. Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons Ltd., Chichester (1994)

[2] Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer Series in Operations Research. Springer-Verlag, New York (1997)

[3] Ben-Tal, A., Nemirovski, A.: Robust optimization: methodology and applications. Math. Program. **92**(3, Ser. B), 453–480 (2002)

[4] Zhang, Y.: General robust-optimization formulation for nonlinear programming. J. Optim. Theory Appl. **132**(1), 111–124 (2007)

[5] Patriksson, M., Wynter, L.: Stochastic mathematical programs with equilibrium constraints. Oper. Res. Lett. **25**(4), 159–167 (1999)

[6] Cromvik, C., Patriksson, M.: On the robustness of global optima and stationary solutions to stochastic mathematical programs with equilibrium constraints, part 2: Applications. J. Optim. Theory Appl. **144**(3) (2010, to appear)

[7] Römisch, W.: Stability of stochastic programming problems. In: Stochastic programming, *Handbooks Oper. Res. Management Sci.*, vol. 10, pp. 483–554. Elsevier, Amsterdam (2003)

[8] Römisch, W., Wets, R.J.B.: Stability of ϵ-approximate solutions to convex stochastic programs. SIAM J. Optim. **18**(3), 961–979 (2007)

[9] Heitsch, H., Römisch, W., Strugarek, C.: Stability of multistage stochastic programs. SIAM J. Optim. **17**(2), 511–525 (2006)

[10] Birbil, Ş.İ., Gürkan, G., Listeş, O.: Solving stochastic mathematical programs with complementarity constraints using simulation. Math. Oper. Res. **31**(4), 739–760 (2006)

[11] Shapiro, A.: Stochastic programming with equilibrium constraints. J. Optim. Theory Appl. **128**(1), 223–243 (2006)

[12] Lin, G.H., Xu, H., Fukushima, M.: Monte Carlo and quasi-Monte Carlo sampling methods for a class of stochastic mathematical programs with equilibrium constraints. Math. Methods Oper. Res. **67**(3), 423–441 (2008)

[13] Evgrafov, A., Patriksson, M., Petersson, J.: Stochastic structural topology optimization: existence of solutions and sensitivity analyses. ZAMM Z. Angew. Math. Mech. **83**(7), 479–492 (2003)

[14] Evgrafov, A., Patriksson, M.: Stochastic structural topology optimization: Discretization and penalty function approach. Struct. and Multidisc. Optim. **25**(3), 17–188 (2003)

[15] Evgrafov, A., Patriksson, M.: Stable relaxations of stochastic stress-constrained weight minimization problems. Struct. Multidiscip. Optim. **25**(3), 189–198 (2003)

[16] Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. Math. Oper. Res. **25**(1), 1–22 (2000)

[17] Fletcher, R., Leyffer, S., Ralph, D., Scholtes, S.: Local convergence of SQP methods for mathematical programs with equilibrium constraints. SIAM J. Optim. **17**(1), 259–286 (2006)

[18] Leyffer, S., López-Calva, G., Nocedal, J.: Interior methods for mathematical programs with complementarity constraints. SIAM J. Optim. **17**(1), 52–77 (2006)

[19] Bard, J.F.: Practical Bilevel Optimization, *Nonconvex Optimization and its Applications*, vol. 30. Kluwer Academic Publishers, Dordrecht (1998)

[20] Luo, Z., Pang, J., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)

[21] Outrata, J., Kočvara, M., Zowe, J.: Nonsmooth approach to optimization problems with equilibrium constraints, *Nonconvex Optimization and its Applications*, vol. 28. Kluwer Academic Publishers, Dordrecht (1998)

[22] Bendsøe, M.P., Sigmund, O.: Topology Optimization. Springer-Verlag, Berlin (2003)

[23] Stackelberg, H.V.: The Theory of Market Economy. Oxford University Press, Oxford (1952)

[24] Nash, J.F.: Non-cooperative games. Ann. of Math. **54**(2), 286–295 (1951)

[25] Evgrafov, A., Patriksson, M.: On the existence of solutions to stochastic mathematical programs with equilibrium constraints. J. Optim. Theory Appl. **121**(1), 65–76 (2004)

[26] Lin, G.H., Chen, X., Fukushima, M.: Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization. Math. Program. **116**(1–2, Ser. B), 343–368 (2009)

[27] Patriksson, M.: On the applicability and solution of bilevel optimization models in transportation science: A study on the existence, stability and computation of optimal solutions to stochastic mathematical programs with equilibrium constraints. Transp. Res. **42B**(10), 843–860 (2008)

[28] Patriksson, M.: Robust bi-level optimization models in transportation science. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **366**(1872), 1989–2004 (2008)

[29] Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. **5**(1), 43–62 (1980)

[30] Clarke, F.H.: Optimization and Nonsmooth Analysis. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons Inc., New York (1983)

[31] Borwein, J.M., Lewis, A.S.: Convex Analysis and Nonlinear Optimization. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 3. Springer-Verlag, New York (2000)

[32] Rockafellar, R., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**(3), 21–41 (2000)

[33] Shapiro, A.: Monte Carlo sampling methods. In: Stochastic Programming, *Handbooks Oper. Res. Management Sci.*, vol. 10, pp. 353–425. Elsevier, Amsterdam (2003)

[34] Shapiro, A., Xu, H.: Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions. J. Math. Anal. Appl. **325**(2), 1390–1399 (2007)

[35] Ye, J.J., Zhu, Q.J.: Multiobjective optimization problem with variational inequality constraints. Math. Program. **96**(1, Ser. A), 139–160 (2003)

[36] Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation. II, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 331. Springer-Verlag, Berlin (2006)

[37] Mordukhovich, B.S.: Multiobjective optimization problems with equilibrium constraints. Math. Program. **117**(1–2, Ser. B), 331–354 (2009)

[38] Ehrgott, M.: Multicriteria Optimization, second edn. Springer-Verlag, Berlin (2005)

[39] Minami, M.: Weak Pareto-optimal necessary conditions in a nondifferentiable multiobjective program on a Banach space. J. Optim. Theory Appl. **41**(3), 451–461 (1983)

[40] Li, X.F.: Constraint qualifications in nonsmooth multiobjective optimization. J. Optim. Theory Appl. **106**(2), 373–398 (2000)

Paper II

On the Robustness of Global Optima and Stationary Solutions to Stochastic Mathematical Programs with Equilibrium Constraints, part 2: Applications

C. Cromvik* M. Patriksson†

Communicated by F. Giannessi

Abstract In a companion paper (Cromvik and Patriksson, On the Robustness of Global Optima and Stationary Solutions to Stochastic Mathematical Programs with Equilibrium Constraints, part 1: Theory, Journal of Optimization Theory and Applications, 2010, to appear) the mathematical modeling framework SMPEC was studied; in particular, global optima and stationary solutions to SMPECs were shown to be robust with respect to the underlying probability distribution under certain assumptions. Further, the framework and theory were elaborated to cover extensions of the upper-level objective: minimization of the conditional value-at-risk (CVaR) and treatment of the multiobjective case. In this paper, we consider two applications of these results: a classic traffic network design problem, where travel costs are uncertain, and the optimization of a treatment plan in intensity modulated radiation therapy, where the machine parameters and the position of the organs are uncertain. Owing to the generality of SMPEC, we can model these two very different applications within the same framework. Our findings illustrate the large potential in utilizing the SMPEC formalism for modeling and analysis purposes; in particular, information from scenarios in the lower-level problem may provide very useful additional insights into a particular application.

*Department of Mathematical Sciences, Chalmers University of Technology and Department of Mathematical Sciences, University of Gothenburg, SE-412 96 Gothenburg, Sweden. E-mail: christoffer.cromvik@chalmers.se

†Corresponding author. Department of Mathematical Sciences, Chalmers University of Technology and Department of Mathematical Sciences, University of Gothenburg, SE-412 96 Gothenburg, Sweden. E-mail: mipat@chalmers.se

Keywords: Stochastic mathematical program with equilibrium constraints, Solution stability and robustness, Traffic network design, Intensity modulated radiation therapy, Sample average approximation

1 Introduction

The framework known as stochastic mathematical program with equilibrium constraints (SMPEC) was introduced in [1] as an extension of the MPEC framework of hierarchical optimization models, in order to incorporate the uncertainty of data that one often must face in applications. Since then, it has found applications in many contexts, such as economics (e.g., [2, 3]), engineering sciences (e.g., [4, 5]), and transportation science (e.g., [6]). The companion paper [7] provides new results on the robustness of solutions to this class of problems. The objective of this paper is to illustrate the potential utilization of these results through two applications. The first is within traffic network design, which is a classic topic (see e.g. [6, 8] and references therein). The second, however, appears to be new: we present an SMPEC model for robust treatment planning in intensity modulated radiation therapy (IMRT), where the machine parameters as well as the position of the organs are uncertain. With this model, we are approaching a robust and in fact deliverable treatment plan.

The remainder of the paper is organized as follows. In Section 2, we summarize briefly the main results from [7]. In Section 3, we provide a small-scale application of the robust design of a traffic network, based on the classic network of Braess. In Section 4, we provide a numerical example of a treatment plan, which accounts for both position uncertainty and the uncertainties in the radiation delivery from a treatment machine.

2 Robustness of Solutions to the SMPEC

Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $y \in \mathbb{R}^m$, $C \subseteq \mathbb{R}^m$ be a polyhedron, let $F(x, \cdot) : C \to \mathbb{R}^m$ be smooth, and let $N_C : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ be the standard normal cone mapping,

$$N_C(y) := \begin{cases} \left\{ z \in \mathbb{R}^m \,|\, z^{\mathrm{T}}(w - y) \leq 0,\ w \in C \right\}, & \text{if } y \in C, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Let $(\Omega, \Theta, \mathrm{P})$ be a complete probability space and consider the problem

$$(\mathrm{SMPEC}_\Omega) \qquad \begin{array}{ll} \min\limits_{(x,y(\cdot))} & \mathrm{E}_\omega[f(x,y(\omega),\omega)] := \displaystyle\int_\Omega f(x,y(\omega),\omega)\, P(d\omega), \\ \text{s.t.} & x \in X, \\ & -F(x,y,\omega) \in N_C(y), \qquad P\text{-a.s.}, \end{array}$$

where $y : \Omega \to \mathbb{R}^m$ is a random element of the probability space $(\Omega, \Theta, \mathrm{P})$. We also introduce $S : \mathbb{R}^n \times \Omega \rightrightarrows \mathbb{R}^m$, which defines the set of solutions to the lower-level parametric variational inequality problem,

$$S(x,\omega) := \{\, y \in \mathbb{R}^m \mid -F(x,y,\omega) \in N_C(y) \,\}.$$

We next summarize briefly the main technical content of [7].

2.1 Stability of Global Solutions and Stationary Points

Assumption A

(A1) The mapping $S(x,\cdot)$ is measurable for any x.

(A2) The set X is closed and the mapping $x \mapsto S(x,\omega)$ is closed for almost any $\omega \in \Omega$.

(A3) The function f is continuous in (x,y), measurable in ω, uniformly weakly coercive with respect to x over the set X, and bounded from below by a (Θ, P)-integrable function.

(A4) The set $S(x_0,\omega)$ is nonempty for some $x_0 \in X$ and almost any $\omega \in \Omega$.

The existence of optimal solutions under Assumption A is established in [9].

Let $\{P_k\}$ be a sequence of probability measures defined on $\mathcal{B}(\Omega)$, and denote by $(\mathrm{SMPEC}_\Omega)^k$ the problem defined by (SMPEC_Ω) with the measure P replaced by P_k.

Theorem 2.1 (Global Stability of Optimal Solutions) *Let Assumption* (A) *hold, suppose that the mapping $F(x,\cdot,\omega)$ is strictly monotone in y for each $x \in X$ and $\omega \in \Omega$, and that the sequence $\{P_k\}$ of probability measures weakly converges to P. Also suppose that, for each k, $(x^k, y^k(\cdot))$ solves* $(\mathrm{SMPEC}_\Omega)^k$. *Then, each limit point (there is at least one) of the sequence $\{(x^k, y^k(\cdot))\}$ is an optimal solution to* (SMPEC_Ω).

Assumptions B

(B1) The function f is Lipschitz continuous in (x, y).

(B2) The mapping $F(\cdot, \cdot, \omega)$ is continuously differentiable and $F(x, \cdot, \omega)$ is uniformly strongly monotone on C for each $x \in X$ and $\omega \in \Omega$.

(B3) $X = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0,\ i = 1, ..., p\}$ and each function g_i is continuously differentiable.

(B4) The Mangasarian–Fromovitz constraint qualification (MFCQ) holds for all $x \in X$.

If Assumptions (B1) and (B2) hold, then ([10]) there exists a locally Lipschitz continuous, single-valued solution map $(x, \omega) \mapsto \sigma(x, \omega)$ with

$$y = \sigma(x, \omega), \qquad \sigma(x, \omega) \in S(x, \omega).$$

With this property, we can rewrite (SMPEC_Ω) as the one-level problem

$$(\mathrm{SNLP}_\Omega) \quad \min_x \quad \mathrm{E}_\omega[f(x, \sigma(x, \omega), \omega)] := \int_\Omega f(x, \sigma(x, \omega), \omega)\, P(d\omega),$$
$$\text{s.t.} \quad x \in X,$$

and correspondingly $(\mathrm{SNLP}_\Omega)^k$ is obtained from (SNLP_Ω) by replacing P with P_k.

Theorem 2.2 (Stability of Stationary Solutions) *Let Assumptions* (A) *and* (B) *hold, suppose that the sequence* $\{P_k\}$ *of probability measures weakly converges to* P *and is upper bounded by a measurable function, and that for each* k, $(x^k, y^k(\cdot))$ *is a Clarke stationary solution to* $(\mathrm{SNLP}_\Omega)^k$. *Then, each limit point (there is at least one) of the sequence* $\{(x^k, y^k(\cdot))\}$ *is a Clarke stationary solution to* (SNLP_Ω).

2.2 Convergence of Sample Average Approximation Schemes

We use the reformulation of $(\mathrm{SMPEC})_\Omega$ into (SNLP_Ω), and consider the problem

$$(\mathrm{SNLP})^N \quad \min_x \quad \hat{f}_N := \frac{1}{N} \sum_{k=1}^{N} f(x, \sigma(x, \omega^k), \omega^k),$$
$$\text{s.t.} \quad x \in X.$$

Assumption C

(C1) The set X is bounded and convex.

(C2) The function $f(\cdot, \sigma(\cdot, \omega), \omega)$ is regular (i.e., f is directionally differentiable and the directional derivative coincides with the Clarke directional derivative) at x for almost any $\omega \in \Omega$.

Theorem 2.3 (Convergence of Optimal Solutions) *Let Assumptions* (A), (B1)–(B2), (C1) *hold. For each N, let $(x_N, y_N(\cdot))$ be an optimal solution to* $(\mathrm{SNLP})^N$. *Then, each limit point (there is at least one) of the sequence $\{x_N\}$ is an optimal solution to* (SNLP_Ω).

Theorem 2.4 (Convergence of Stationary Solutions) *Let Assumptions* (A), (B), (C) *hold. For each N, let $(x_N, y_N(\cdot))$ be a stationary solution to* $(\mathrm{SNLP})^N$. *Then, each limit point (there is at least one) of the sequence $\{x_N\}$ is a stationary solution to* (SNLP_Ω).

Since (SNLP_Ω) is a reformulation of (SMPEC_Ω), the above theorems state that we also have convergence to global optima and stationary solutions for the corresponding discretized problem

$$
\begin{array}{lll}
(\mathrm{SMPEC})^N & \min\limits_{x} & \hat{f}_N := \dfrac{1}{N}\displaystyle\sum_{k=1}^{N} f(x, y^k, \omega^k), \\
& \text{s.t.} & x \in X, \\
& & -F(x, y^k, \omega^k) \in N_C(y^k), \qquad k = 1, \ldots, N.
\end{array}
$$

2.3 Extensions to a Risk Objective Function and to Multiple Objectives

We consider next the SMPEC where the expected value in the objective function is replaced by an expression for conditional value-at-risk (CVaR) at level β

$$
\begin{array}{lll}
(\mathrm{SRPEC})_\Omega & \min\limits_{(x, y(\cdot), \gamma)} & \gamma + \dfrac{1}{1-\beta}\displaystyle\int_\Omega [f(x, y(\omega), \omega) - \gamma]_+ \, P(d\omega), \\
& \text{s.t.} & x \in X, \\
& & -F(x, y, \omega) \in N_C(y), \qquad P\text{-a.s.},
\end{array}
$$

and analogously for the problem $(\mathrm{SRPEC}_\Omega)^k$. In line with Theorem 2.1, we can establish robustness of global optima; by reformulating SMPEC into

one-level problems,

$$(\mathrm{SRNLP}_\Omega) \quad \min_{(x,\gamma)} \quad \gamma + \frac{1}{1-\beta} \int_\Omega [f(x, \sigma(x,\omega), \omega) - \gamma]_+ \, P(d\omega),$$
$$\text{s.t.} \quad x \in X,$$

we can also establish robustness of stationary solutions as in Theorem 2.2, and the convergence of the SAA scheme as in Theorems 2.3 and 2.4; see [7] for details.

We define the multiple objective version of the SMPEC, the SMOPEC, for q objectives as that to

$$(\mathrm{SMOPEC}_\Omega) \quad \min_{(x,y(\cdot))} \quad (\mathrm{E}_\omega[f_1(x, y(\omega), \omega)], \ldots, \mathrm{E}_\omega[f_q(x, y(\omega), \omega)]) \,,$$
$$\text{s.t.} \quad x \in X,$$
$$-F(x, y, \omega) \in N_C(y), \qquad \text{P-a.s.},$$

and correspondingly $(\mathrm{SMOPEC}_\Omega)^k$ is obtained by replacing P with P_k.

Theorem 2.5 (Stability of Weakly Pareto-Optimal Solutions to (SMOPEC_Ω)) *Let Assumption* (A) *hold, suppose that the mapping $F(x, \cdot, \omega)$ is strictly monotone for each $x \in X$ and $\omega \in \Omega$, that* (SMOPEC_Ω) *is a convex problem, and that the sequence $\{P_k\}$ of probability measures weakly converges to P. Also, suppose that, for each k, $(x^k, y_k(\cdot))$ is a weakly Pareto-optimal solution to* $(\mathrm{SMOPEC}_\Omega)^k$. *Then, each limit point (there is at least one) of the sequence $\{(x^k, y^k(\cdot))\}$ is a weakly Pareto-optimal solution to* (SMOPEC_Ω).

To establish stability without a convexity assumption, we reformulate (SMOPEC_Ω) and $(\mathrm{SMOPEC}_\Omega)^k$ as one-level problems by treating y as a function of x and ω, $y = \sigma(x, \omega)$. This is possible if, in addition to the assumptions in Theorem 2.5, Assumptions (B1) and (B2) hold. We denote the reformulations by (SMONLP_Ω) and $(\mathrm{SMONLP}_\Omega)^k$, respectively, where the first problem has the following appearance:

$$(\mathrm{SMONLP}_\Omega) \quad \min_{x} \quad (\mathrm{E}_\omega[f_1(x, \sigma(x,\omega), \omega)], \ldots, \mathrm{E}_\omega[f_q(x, \sigma(x,\omega), \omega)]),$$
$$\text{s.t.} \quad x \in X.$$

Theorem 2.6
(Stability of Weakly Pareto-Stationary Solutions to (SMONLP_Ω))
Let Assumptions (A) *and* (B) *hold, suppose that the sequence $\{P_k\}$ of probability measures weakly converges to P, and that, for each k, $(x^k, y^k(\cdot))$ is a weakly Pareto-stationary solution to* $(\mathrm{SMONLP}_\Omega)^k$. *Then, each limit point (there is at least one) of the sequence $\{(x^k, y^k(\cdot))\}$ is a weakly Pareto-stationary solution to* (SMONLP_Ω).

3 First Application: Traffic Network Design

We consider a road traffic model. The network is represented by a strongly connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of directed links. For each origin–destination (OD) pair $(p, q) \in \mathcal{C}$ with $\mathcal{C} \subset \mathcal{V} \times \mathcal{V}$, there is a transportation demand. Each route $r \in \mathcal{R}_{pq}$ joining the OD pair (p, q) has an associated flow h_r and a travel cost c_r.

We assume that the design parameter $x \in \mathbb{R}^n$ influences the travel cost and the demand. The travel cost function has the form $c(x, \cdot) : \mathbb{R}_+^{|\mathcal{R}|} \to \mathbb{R}^{|\mathcal{R}|}$, where $|\mathcal{R}|$ is the total number of routes. The demand for each OD pair depends on the travel cost and the demand function has the form $d(x, \cdot) : \mathbb{R}^{|\mathcal{C}|} \to \mathbb{R}_+^{|\mathcal{C}|}$.

Wardrop's user equilibrium condition [11] states that, for each OD pair, the travel cost for all routes utilized must be equal and minimal. Since the flow is nonnegative, this condition can be formulated as a complementarity condition. Let π_{pq} be the minimum travel cost for the OD pair (p, q). The equilibrium condition is that

$$0 \leq h_r \perp c_r(x, h) - \pi_{pq} \geq 0, \qquad r \in \mathcal{R}_{pq}, \qquad (p, q) \in \mathcal{C}, \tag{1}$$

where $a \perp b$ denotes the condition $a^{\mathrm{T}} b = 0$. Utilizing a route–OD pair incidence matrix $\Gamma \in \mathbb{R}^{|\mathcal{R}| \times |\mathcal{C}|}$, we can express demand feasibility as follows:

$$\Gamma^{\mathrm{T}} h = d(x, \pi). \tag{2}$$

Combining (1) and (2), we characterize the user equilibrium flows as a mixed complementarity problem (MCP),

$$0^{|\mathcal{R}|} \leq h \perp c(x, h) - \Gamma\pi \geq 0^{|\mathcal{R}|}, \tag{3a}$$

$$\Gamma^{\mathrm{T}} h = d(x, \pi). \tag{3b}$$

If we assume that the travel cost is positive, then (3) can instead be formulated as the following nonlinear complementarity problem (NCP) ([12, 13]):

$$0^{|\mathcal{R}|} \leq h \perp c(x, h) - \Gamma\pi \geq 0^{|\mathcal{R}|},$$
$$0^{|\mathcal{C}|} \leq \pi \perp \Gamma^{\mathrm{T}} h - d(x, \pi) \geq 0^{|\mathcal{C}|}.$$

We can also provide a link flow representation of the user equilibrium condition. Let $v \in \mathbb{R}^{|\mathcal{E}|}$ be a vector of link flows and let $t_l(x, v)$ be the link travel cost for $l \in \mathcal{E}$. If we assume that the travel costs are additive and introduce a route–link incidence matrix $\Lambda \in \{0, 1\}^{|\mathcal{E}| \times |\mathcal{R}|}$, the link travel cost is related to the route travel cost through the relation $c(x, h) = \Lambda^{\mathrm{T}} t(x, v)$. Also, to have flow conservation, we require that

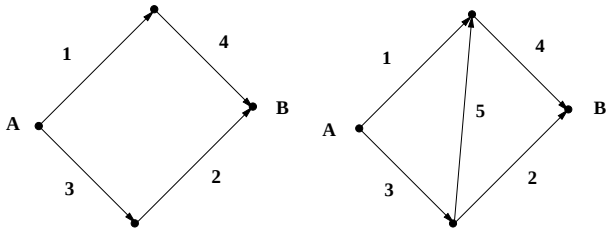
$$v = \Lambda h. \tag{4}$$

The main objective in a network design problem is to influence the travel costs and the demands such that some criterion is optimized. The design problem can be formulated as an MPEC, where the traffic equilibrium is described by the system (3) of mixed complementarity constraints. An example of a network design problem is given by setting link tolls through the design parameter $x \in \mathbb{R}^n$, with $n \leq |\mathcal{E}|$, such that the total travel cost $f(x,v) := \sum_{l \in \mathcal{E}} t_l(x,v)v_l$ is minimized, and where, for a given design x, v is given by (3) and (4).

The traffic equilibrium model is a static model. All quantities are assumed to be an average over a time period, and as such they are subjected to uncertainties. The travel costs $t(x,v,\omega)$ and demands $d(x,\pi,\omega)$ are to some extent uncertain and can change depending on external factors, such as the weather. Therefore, we can and should formulate the network design problem as an SMPEC, which gives us a design which is the best possible on average. This problem has been studied in Patriksson [6, 8]. Birbil et al. [14] consider a similar model in which, however, the response variables are not stochastic, but are solutions to a stochastic equilibrium problem.

For further references on traffic equilibrium models, see [13, 15, 16].

We present a small numerical example in the application of network design under user equilibrium. The deterministic example is known as Braess' paradox (see e.g. [15, page 75]). It demonstrates that adding an extra link to a network can cause an increase in the total travel cost. In short, this is due to the fact that user equilibrium is a selfish optimum and not a system optimum. Figures 1(a) and 1(b) show the network graph with four and five links, which we will refer to as graph I and graph II, respectively.



(a) Original network.

(b) Network with the addition of a new link.

Figure 1: Networks I and II.

We have one OD-pair (A,B) with a fixed demand of $d = 6$ units. The original network has two paths, using the links $(1,4)$ and $(3,2)$, respectively;

network II has three paths, using the links $(1,4)$, $(3,2)$ and $(3,5,4)$, respectively. The link travel costs are $t_i = 50+v_i$ for $i = 1,2$, $t_i = 10v_i$ for $i = 3,4$, and $t_5 = 10+v_5$. Given theses costs, the user equilibrium flows for network I are $v = (3,3,3,3)^{\mathrm{T}}$, $h = (3,3)^{\mathrm{T}}$. These flows give the equilibrium travel cost $\pi = 83$. For network II, the user equilibrium flows are $v = (2,2,2,2,2)^{\mathrm{T}}$, $h = (2,2,2)^{\mathrm{T}}$. These flows give the equilibrium travel cost $\pi = 92$. Note that adding a link to network I yields an increase in the equilibrium travel cost.

The idea is to set tolls on network II such that we minimize the total travel cost $T(x,v) = \sum_{i=1}^{5} t_i v_i$. For the example above, we consider adding a toll x on the new link, thus altering the travel cost to $t_5 = 10 + v_5 + x$, and consider the problem

$$\begin{aligned} \min_{(x,v,\pi,h)} \quad & T(x,v) + \tau x^2, \\ \text{s.t.} \quad & x \in X, \\ & (v,\pi,h) \text{ solves (3) and (4)}, \end{aligned}$$

where $X = \{x \in \mathbb{R} \mid 0 \le x \le 14\}$ and $\tau > 0$ is a penalty parameter against setting a too high toll value. For a sufficiently small value of τ, the optimal solution is $x^* = 13$ and the optimal total travel cost is $T(x^*, v^*) = 498$. The optimal solution $x^* = 13$ is the threshold value for which there will be no flow on link 5, which in turn will give a lower total travel cost.

Now, consider the case when the travel costs are stochastic. In particular, let us assume that the travel costs on links 3 and 4 are

$$t_i = 10v_i + \omega_{i-2}, \quad i = 3,4, \tag{5}$$

and that each component in ω is independent and drawn from a normal distribution with mean 0 and variance 1, i.e., $\omega \sim \mathcal{N}(0, \mathrm{diag}(1,1))$. We consider the following SMPEC model:

$$\begin{aligned} \min_{(x,v(\cdot),\pi(\cdot),h(\cdot))} \quad & \mathrm{E}_\omega[f(x,v(\omega))] := \mathrm{E}_\omega[T(x,v(\omega))] + \tau x^2, \\ \text{s.t.} \quad & x \in X, \\ & 0 \le h(\omega) \perp \Lambda^{\mathrm{T}} t(x,v(\omega)) - \Gamma\pi(\omega) \ge 0, && P\text{-a.s.}, \\ & \Gamma^{\mathrm{T}} h(\omega) = d, && P\text{-a.s.}, \\ & v(\omega) = \Lambda h(\omega), && P\text{-a.s.}, \end{aligned}$$

where Λ and Γ are the route–link incidence matrix and the route–OD pair incidence matrix, respectively, for network II. The SMPEC is solved using the discretization scheme SAA. Since the travel costs $t(x,\cdot)$ are strongly monotone, the SMPEC satisfies Assumptions A–C, and this implies that

the optimal solution is stable in the sense of Theorems 2.1 and 2.2 (see also [8]) and that SAA converges by Theorems 2.3 and 2.4.

Note that the number of variables and constraints scale linearly with the number of scenarios, since we use a general nonlinear optimization solver. This means that, for 100 scenarios, the problem has 900 variables. So, even though the deterministic problem is of small scale, the stochastic version is of large scale. By switching to an implicit function $v(x, \omega)$ and using sensitivity analysis ([17]), we may be able to solve larger problems. The traffic model was implemented in MATLAB and solved using the solver SNOPT [18].

For a run with $N = 400$, the solver converged to the stationary solution $x^* = 14$. In Figure 2, we plot histograms of the objective values for stationary solutions to three models: one with the expected value in the objective ($x^* = 14.0$), one with CVaR at $\beta = 0.8$ ($x^* = 11.6$) and one with CVaR at $\beta = 0.95$ ($x^* = 10.6$).

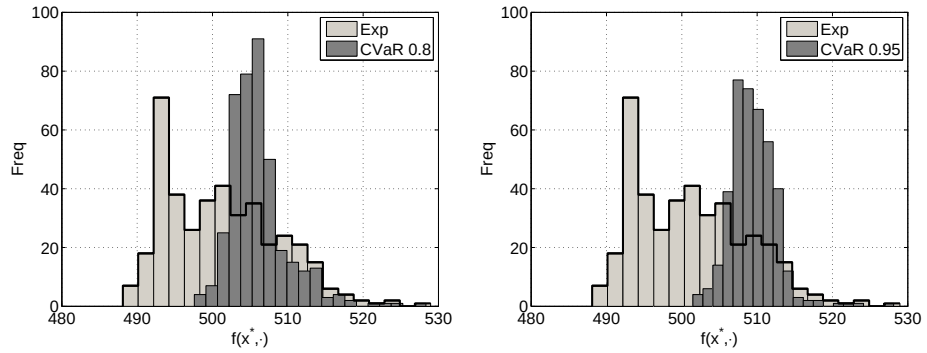


Figure 2: Histograms ($N = 400$) for the objective values: one with expected value in the objective, one with CVaR at $\beta = 0.8$, and one with CVaR at $\beta = 0.95$.

In order to illustrate the influence of the variance of the uncertain parameter on the solution, we show in Figure 3 histograms of the equilibrium path travel cost for stationary solutions corresponding to four values of the variance of the stochastic variable. The results are not surprising: a larger variance implies a larger spread in the response. (We note in passing that Gwinner and Raciti [19] consider the stochastic traffic network equilibrium model, i.e., the lower-level problem in the SMPEC model, and have developed a procedure for the analytical computation of the mean equilibrium flows and their variance for the case when the travel costs are affine in the flow variables.) Having access to histograms for responses, i.e., equilibrium solutions, is a feature of SMPEC which may be valuable for getting specific

insights into an application.

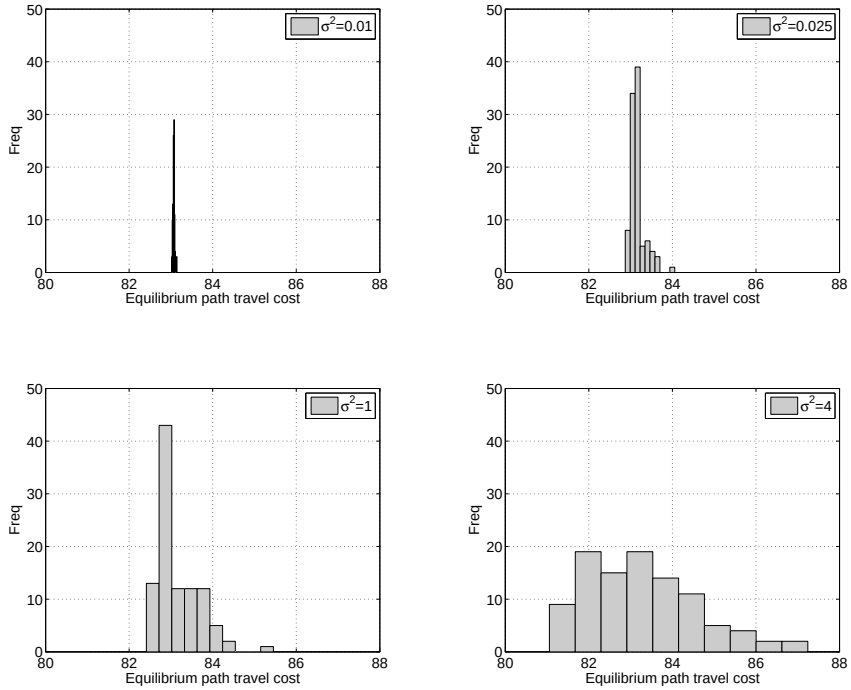


Figure 3: Equilibrium path travel costs for runs with $N = 100$ and with variance $\sigma_j^2 = 0.01$, $\sigma_j^2 = 0.025$, $\sigma_j^2 = 1$, and $\sigma_j^2 = 4$, $j = 1, 2$, for the link costs in (5). Each subfigure corresponds to one stationary solution.

Regarding the solution of the discretized model, we have made the natural observation that the solution time increases with the number of scenarios and that it also increases with the variance of the stochastic variable. Obviously, these observations are based on a single test case and should not be considered as general conclusions.

The design and implementation of a pricing scheme must be simple and transparent; it must also address social welfare issues such as the welfare effect of tolls across population groups. In [20–22], several equity measures are presented and evaluated in the context of optimal network design. In the first two papers, these design models are built upon stochastic (in fact, probit) traffic equilibrium models; such equilibrium solutions are still deterministic functions of the data of the traffic network. The results of our accompanying paper [7], as outlined in the previous section, are immediately transferable to such a setting; see also the discussion sections in [6, 8].

Regardless of whether the traffic model is a deterministic or a stochastic equilibrium one, the theoretical results of our accompanying paper [7] may provide ideas for several interesting developments. First, we may theoretically validate and numerically study SMPEC versions of network design problems where equity is included as an objective. In the previous papers [20–22] equity measures are treated through upper-level constraints, which however in general result in unstable optimal and stationary solutions. Our proposal is to instead study a bicriterion version of the problem (see the early reference [23] on multiobjective traffic network design), where the original toll efficiency objective is complemented by an equity objective. Moreover, as network user responses in the SMPEC model are stochastic, the equity objective is natural to include as a risk measure.

While link tolling is a decentralized pricing mechanism, signal timings are centralized mechanisms for controlling traffic flows; optimizing signal controls subject to traffic equilibrium constraints is nearly as old a scientific subject as is toll optimization; see e.g. the reviews in [24, 25]. It is frequently recognized that the performance of signal timings is unstable due to fluctuating traffic conditions, such as fluctuating demands (see e.g. [26]). The above example serves to illustrate the potential in utilizing the SMPEC formalism in this field.

4 Second Application: Optimization of a Treatment Plan for Radiation Therapy

In radiation therapy, cancerous tumors are subjected to ionizing radiation. The objective is to eradicate the tumor while sparing the surrounding tissue and organs at risk. We will show that MPEC models can be utilized to find optimal radiation plans.

Radiation is delivered by a linear accelerator and, by using what is called multileaf collimators, the radiation beam can be shaped such that different parts in the treatment region receive different doses. This technique of shaping the beam is called intensity modulated radiation therapy (IMRT). Since there are millions of ways of modulating the intensity, the most suitable radiation dose is found by optimization. The ideal dose is still often not attainable, so the objective is to find the best compromise achievable.

The linear accelerator can deliver radiation to the target from several angles by the use of a gantry arm. For some cases, up to nine gantry angles are used to give a good target coverage. The angles are usually considered fixed in the optimization problem.

We will describe two methods for parameterizing the multileaf collimator system. Both methods may act as the lower-level problem in an (S)MPEC

setting. The beam cross-section is subdivided into small rectangular cells, known as beamlets, and the decision variables are the intensities through each cell. In the first method, we assume that any bounded, nonnegative, intensity profile is attainable by multileaf collimators as a total over a treatment time. We also assume that the doses scale linearly with the intensities and are additive.

In the other method, the leaf trajectories are parameterized; given a desired intensity, the trajectories are found through an optimization problem.

Objective functions in IMRT are either physically or biologically based. In short, a physically based function is a function of the dose alone. It can, for example, be the quadratic deviation from a dose level sought or the maximum dose in a domain. A biologically based function is associated with a specific organ and measures the biological effect of a dose. The function is constructed using a set of biological parameters which can depend on the organ type, its size, shape etc.

An example of a biologically based function is the normal tissue complication probability (NTCP) (see e.g. [27]):

$$\text{NTCP} = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \exp^{-t^2/2}\, dt,$$

where

$$u = \frac{\text{GEUD}(d) - D_{50}}{mD_{50}},$$

and D_{50} is the homogeneous dose corresponding to 50% risk of complication, m determines the slope of the risk, d is the dose, and GEUD ([28]) is the generalized equivalent uniform dose,

$$\text{GEUD}(d) = \left(\tfrac{1}{|J|} \sum_{j \in J} d_i^a \right)^{1/a},$$

where J is the set of voxels (discretized cells) in the organ. The parameter a influences the volume effect of the dose: if $a = 1$, then the function measures the mean dose; for higher values, the function value approaches the maximum dose. For example, the spinal cord is sensitive to a maximum dose and the volume effect is low. On the other hand, the parotid glands are organs that are sensitive to how much volume receives a certain dose. Suitable values for a are typically fit from clinical data. The GEUD function may be used as an objective function by itself. The function is attractive from several points of view, not the least the fact that it is convex if $a \geq 1$, which is the case for organs.

Radiation therapy is delivered in fractions over several weeks; although the patient is fixated, there will be variations in position over the sessions. (This is called setup errors.) Another uncertainty that affects how good a treatment plan is in practice is the patient and organ motion which will vary during a treatment session. Combining the two, we get a position uncertainty. Olafsson and Wright [29] and Chu et al. [30] assume that the doses are stochastic and use probabilistic constraints to control the dose levels in the target and in the organs at risk. Chan et al. [31] use a motion probability mass function and assumes that the probability itself is uncertain. Using linear programming duality, they can formulate the optimization problem as a large linear program. Baum et al. [32] use coverage probabilities for the target and tumor as penalties in the objective functions to derive a robust treatment. Unkelbach and Oelfke [33] discuss, from a mathematical and a physics perspective, the difference between using coverage probabilities and stochastic programming in IMRT optimization.

Biological uncertainty can also be incorporated in an SMPEC model through the objective function. An optimal solution then is the best from a population perspective. Functions based on the biological effect have certain advantages over physically based ones, but they rely on the accuracy of biological parameters which are fit from data in medical studies. For example, the dose–volume effect for the bladder is uncertain (see e.g. [34, 35]), which has an impact on the parameter a in the GEUD function.

In Kåver et al. [36] and Lian and Xing [37], stochastic programming is used for the optimization of a treatment plan when there are uncertainties in the biological parameters. In both papers, the expected value of the objective functions is minimized. Kåver et al. use the objective P_+ ([38]), which is a nonconvex objective function; Lian and Xing use an objective function based on GEUD. Lian and Xing report that the result strongly depends on the underlying probability distribution.

For more general information on IMRT, see e.g. [39–41].

4.1 Linear Lower-Level Problem

Let $y \in \mathbb{R}^m$ denote the dose in the voxels and let $x \in X$ denote the intensities (beamlets) in the beam. If the dose scales linearly and is additive, an attainable dose is given by the equation

$$y = Kx, \tag{6}$$

where the influence matrix $K \in \mathbb{R}^{m \times n}$ is computed beforehand. To put this parameterization into the (S)MPEC framework, let x denote the decision variables and let y denote the response variables. The equation (6) then represents the lower-level problem.

In connection with this setting, we consider a prostate case where the PTV (Planned Target Volume) overlaps two critical structures: the rectum and the bladder (see Figure 4). We have four objectives, which are listed in Table 1. The PTV should receive a uniform dose of $d^P = 70$ Gy (J/kg). The rectum is considered as an organ with a serial architecture, which means that it is sensitive to the maximum dose. The volume effect parameter for this organ is set to $a_r = 8.3$ according to Emami et al. [42]. The architecture of the bladder is more uncertain, but it is set to $a_b = 2$ [42].

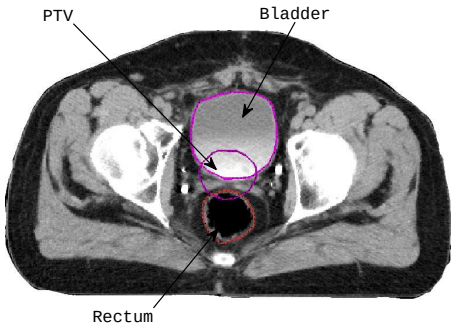


Figure 4: Transverse CT scan.

In Table 1, the structure 'Unspecified' refers to the normal tissue surrounding the other structures. We enforce a maximum dose limit $d^N = 50$ Gy on this structure. This is quite common in clinical practice in order to avoid hot spots which can induce secondary cancers.

Structure	Prescription
PTV	Uniform dose of $d^P = 70$ Gy ($\pm 5\%$)
Bladder	gEUD ($a = 2$) below 32 Gy
Rectum	gEUD ($a = 8.3$) below 58 Gy
Unspecified	Maximum dose of $d^N = 50$ Gy

Table 1: Treatment parameters.

The upper limits for the GEUD for the rectum g_r and for the bladder g_b were computed from the treatment plan which was used in the clinic for this patient.

Since the structures overlap, the four objectives will be in conflict with each other; it is our goal to find a good compromise. This is a multiple objective problem for which Theorem 2.5 and 2.6 represent the stability results obtainable. We use a simple scalarization of the objective functions.

Target coverage is enforced by minimizing a quadratic measure of the deviation. We assume that the location of the tumor is uncertain and consider it to move like a rigid object with a radial offset. We choose to ignore any fractionation effect and consider the treatment to be given at a single session. This means that we assume that the total dose D for the target after N fractions is $D(\omega) = \sum_{i=1}^{N} d(\omega^i)$; it is assumed to have its center normally distributed with standard deviation 0.6 cm. Furthermore, let $\mathcal{T}(\omega)$, $\mathcal{B}$, $\mathcal{R}$, $\mathcal{N}$ denote the voxels in the PTV, the bladder, the rectum and the normal tissue, respectively. For a vector $v \in \mathbb{R}^n$ and a set $I = \{i_1, \ldots, i_k\}$ with $|I| \leq n$, we use the notation

$$v_I = (v_{i_1}, \ldots, v_{i_k})^{\mathrm{T}} \quad \text{and} \quad v_+ = (\max\{0, v_1\}, \ldots, \max\{0, v_n\})^{\mathrm{T}}.$$

Also, we let $e = (1, \ldots, 1)^{\mathrm{T}}$.

Consider the following multiobjective problem, which is of the form (SMOPEC$_\Omega$):

$$\begin{array}{lll} (\mathrm{SIMRT}_\Omega) & \min\limits_{(x,d)} & (\mathrm{E}_\omega[Q_1(d,\omega)], Q_2(d)), \\ & \text{s.t.} & x_j \in X_j, \quad j = 1, \ldots, n, \\ & & d = Kx, \end{array}$$

where $X_j = [0, u_j]$ are lower and upper bounds on the intensities and the objective functions are defined as

$$\begin{aligned} Q_1(d,\omega) &:= \tfrac{1}{2|\mathcal{T}(\omega)|}(d_{\mathcal{T}(\omega)} - s_0)^{\mathrm{T}} S_0^{-1}(d_{\mathcal{T}(\omega)} - s_0), \\ Q_2(d) &:= \tfrac{1}{g_r}\left(\tfrac{1}{|\mathcal{R}|}\sum_{j\in\mathcal{R}} d_j^{a_r}\right)^{1/a_r} + \tfrac{1}{g_b}\left(\tfrac{1}{|\mathcal{B}|}\sum_{j\in\mathcal{B}} d_j^{a_b}\right)^{1/a_b} \\ &\quad + \tfrac{1}{2|\mathcal{N}|}(d_{\mathcal{N}} - r_0)_+^{\mathrm{T}} R_0^{-1}(d_{\mathcal{N}} - r_0)_+, \end{aligned}$$

where

$$s_0 = d^P e,\ S_0 = \mathrm{diag}(s_0),\ r_0 = d^N e,\ R_0 = \mathrm{diag}(r_0).$$

Both Q_1 and Q_2 are convex functions; hence, (SIMRT$_\Omega$) is a convex problem. The function Q_1 measures the deviation from the target dose d^P; Q_2 is based on the GEUD functions for the rectum and the bladder, respectively, and a one-sided measure of the deviation from the maximum dose in the unspecified tissue.

The uncertainty in (SIMRT$_\Omega$) enters only into the objective functions. The lower-level problem is the trivial linear equation, $d = Kx$, and so the assumptions of Theorem 2.5 are fulfilled: weakly Pareto-optimal solutions are stable.

We use beams from 5 equidistant gantry angles to irradiate the tumor. The number of voxels and variables in (SIMRT_Ω) depends on the number of beams, the beamlet size and the geometry and resolution of the patient region. For this case, we have $1,526,330$ voxels and $n = 336$ variables. We let $u_j = 30$ for $j = 1, \ldots, n$. The IMRT model was implemented in Fortran 90. The radiation treatment planning tool CERR [43] was used to setup the problem, and LANCELOT B [44] was used as the optimization solver.

We compare three models: a conventional treatment where we use a static target with a 1 cm extra margin ($\mathcal{T}(\omega)$ in (SIMRT_Ω) is replaced by a static set); a treatment with (SIMRT_Ω); and a treatment with (SIMRT_Ω) but with conditional value-at-risk at level $\beta = 0.8$ instead of the expected value. We compute only one solution on the Pareto surface by setting the constraint $Q_1 \leq 0$; see Figure 5. Using an extra margin is the most conservative choice as it will "guarantee" that the tumor gets a sufficient dose, although at the expense of extra radiation to the risk organs. The expected value is the least conservative choice, as it constrains the tumor in a mean sense; the CVaR objective is a compromise. The conventional treatment gives a 2.7% risk of rectal bleeding ($m = 0.15, D_{50} = 80$ Gy [42]), while the expected value results in the risk being reduced to 0.9%. If CVaR at level $\beta = 0.8$ is used, the risk is 1.1%.

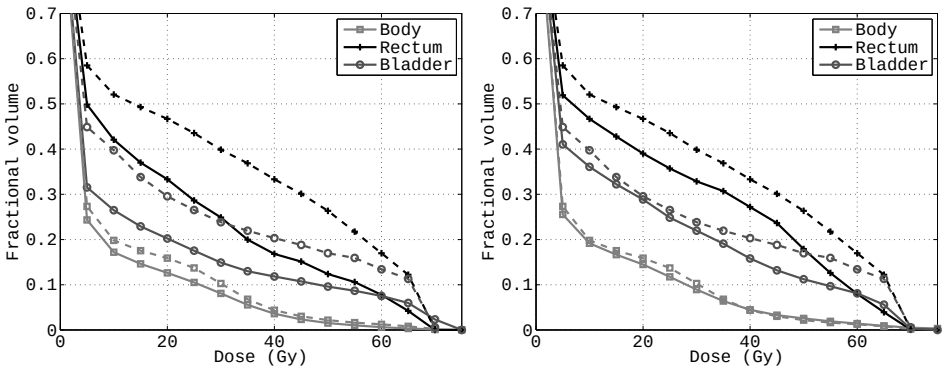


Figure 5: Dose-volume histograms for the organs at risk: bladder, rectum, and normal tissue. The dashed line corresponds to the conventional treatment (both figures), while the solid line corresponds to the expected value objective (left figure) and the CVaR objective (right figure).

The results show that, if there is a willingness to be less conservative in the choice of target coverage, then there is a benefit in using stochastic models in terms of risk of complications.

4.2 Deliverable Treatment

The multileaf collimators (MLC) that shape the beam are organized as pairs. If we consider the rectangular beam cross-section as a matrix, where each element represents a discretized cell, then each leaf pair can block a "column"; see Figure 6 (note that, in the figure, each leaf blocks one row). Let $I = \{1, \ldots, m\}$, $J = \{1, \ldots, n\}$, and let x_{ij}, $j \in J$, $i \in I$, be the desired intensity (decision variables) in beamlet (i, j). We assume that there are n pairs of MLC leaves (A_j, B_j), $j \in J$. We assume that the leaves move from row 1 to row m.

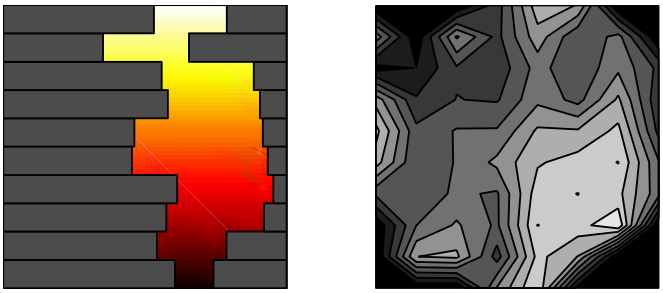


Figure 6: Snapshot of the motion of the multileaf collimators (left figure) for the fluence (intensity) profile shown in the right figure for a cross-section of the beam.

Let a_{ij} and b_{ij} denote the cumulative beam-on time in monitor units of leaf A_j and leaf B_j, respectively, at row i. Assuming that each leaf totally blocks the radiation, and that it "jumps" instantaneously from row i to row $i + 1$, the delivered intensity y satisfies $y_{ij} = a_{ij} - b_{ij}$. The objective with leaf motion computation is to determine the beam-on time a and b such that the desired intensity is obtained, in minimum total time (beam-on time).

The leaves are constrained to not exceed a maximum speed, which implies that there is a minimum time difference between the beam-on time for two consecutive rows. If we assume that the distance between all rows are constant, we can set the minimum time difference to Δt.

We assume that each pair of leaves begin at row 1. In reality, leaf B could be positioned at $i > 1$ if the intensities permit it. This would reduce the total time.

Since the beam must be on until the slowest leaf pair finishes, the leaves A_j are also constrained to end at row m at the same time. This would avoid any extra radiation through the unclosed leaves. The optimization problem

to deliver an intensity x, in minimum time, is formulated as:

$$
\begin{array}{lll}
\text{(TLP)} \quad \min\limits_{(t,a,b)} & t, & \\
\text{s.t.} & a_{i+1j} \geq a_{ij} + \Delta t, & i = 1, \ldots, m-1, \\
& b_{i+1j} \geq b_{ij} + \Delta t, & i = 1, \ldots, m-1, \\
& a_{m,j} = t, & j \in J, \\
& x_{ij} = a_{ij} - b_{ij}, & i \in I, j \in J, \\
& a_{ij} \geq 0, & i \in I,\ j \in J, \\
& b_{ij} \geq 0, & i \in I,\ j \in J.
\end{array}
$$

This model was formulated by Convery and Rosebloom [45]. Spirou and Chui [46] present an analytic expression for the optimal solution, which is derived from the fact that one of the leaves in each pair must move at the maximum speed. Consider the pair j and assume that a_{ij} and b_{ij} are known.

$$
\text{If } x_{i+1j} \geq x_{ij}, \text{ then} \qquad \begin{cases} b_{i+1j} = b_{ij} + \Delta t, \\ a_{i+1j} = b_{i+1j} + x_{i+1j}, \end{cases} \tag{7a}
$$

$$
\text{If } x_{i+1j} < x_{ij}, \text{ then} \qquad \begin{cases} a_{i+1j} = a_{ij} + \Delta t, \\ b_{i+1j} = a_{i+1j} - x_{i+1j}. \end{cases} \tag{7b}
$$

In reality, there are phenomena which effect the delivered intensity y. One such is leaf transmission, which is considered by Spirou and Chui [46]. Let τ be the leaf transmission factor and let t^* be the total beam-on time. Then the intensity delivered is

$$
y_{ij} = \tau[t^* - (a_{ij} - b_{ij})] + a_{ij} - b_{ij}, \tag{8}
$$

where $t^* - (a_{ij} - b_{ij})$ is the total time cell (i, j) is blocked and there is a transmission. From the analytic expressions (7), the total beam-on time is given by ([46])

$$
t^* = \max_j \left(n\Delta t + x_{mj} + \sum_{i=1}^{m-1} [x_{i+1j} - x_{ij}]_+ \right).
$$

We consider the linear optimization problem (TLP) as the lower-level problem in an (S)MPEC framework. We let x denote the decision variables. If we let the total beam-on time t denote the response variable (a and b are eliminated in (8) by the equation $x_{ij} = a_{ij} - b_{ij}$), we see that, since it is Lipschitz continuous, it fulfills the requirements on the lower-level problem for global and stationary stability, cf. Theorem 2.1 and 2.2. On the other

hand, if we let a and b denote the response variables, we note that they are not Lipschitz continuous with respect to x as is shown in the analytic expressions. However, in practice, if we consider a nonoptimal heuristic solution in the lower-level problem, we can modify the analytic expressions with a ramp to make the responses A and B Lipschitz continuous with respect to x, and this will give stability.

We now consider the same prostate case as in the previous subsection, with the exception that the dose to the tumor should be $\pm 3\%$. We assume that the locations are static, but the leaf trajectories A and B are stochastic. This is manifested in that, at each position (i,j), there can be a delay in the beam-on time, which corresponds to the fact that the leaves may unexpectedly move more slowly than anticipated. This is to some extent observed in reality ([47]). Consider the following stochastic multiobjective problem:

$$
\begin{aligned}
(\mathrm{SDEL}_\Omega) \qquad \min_{(x,y(\cdot),a(\cdot),b(\cdot),t(\cdot))} \; & (\mathrm{E}_\omega[G_1(d(\omega))], \mathrm{E}_\omega[G_2(d(\omega))]) \,, \\
\text{s.t.} \quad & x \in X, \\
& y(\omega) = \tau(t-(a-b)) \\
& \qquad + (1+c(\omega))(a-b), && P\text{-a.s.}, \\
& d(\omega) = Ky(\omega), && P\text{-a.s.}, \\
& (a,b,t) \text{ solves (TLP) given } x,
\end{aligned}
$$

where c determines the speed lag and

$$
\begin{aligned}
G_1(d(\omega)) &:= \tfrac{1}{2|\mathcal{T}|}(d_{\mathcal{T}} - s_0)^{\mathrm{T}} S_0^{-1}(d_{\mathcal{T}} - s_0), \\
G_2(d(\omega)) &:= \tfrac{1}{g_r}\left(\tfrac{1}{|\mathcal{R}|}\sum_{j\in\mathcal{R}} d_j^{a_r}\right)^{1/a_r} + \tfrac{1}{g_b}\left(\tfrac{1}{|\mathcal{B}|}\sum_{j\in\mathcal{B}} d_j^{a_b}\right)^{1/a_b} \\
&\quad + \tfrac{1}{2|\mathcal{N}|}(d_{\mathcal{N}} - r_0)_+^{\mathrm{T}} R_0^{-1}(d_{\mathcal{N}} - r_0)_+.
\end{aligned}
$$

The function c is chosen such that, for each beam, 10% of the positions are affected. Let $k \in \mathcal{K}$ denote the affected intensities, let $\omega_k \sim \mathcal{N}(0.1, 0.1) \cap \mathbb{R}_+$ for $k \in \mathcal{K}$, let $c_k = \omega_k$ for $k \in K$, and let $c_k = 0$ for $k \notin K$. The multiobjective problem is discretized using SAA with 20 samples and solved using the epsilon-constrained method ([48]). The model was implemented in Fortran 90 and LANCELOT B [44] was used as the optimization solver. The lower-level problem was implemented as an implicit function.

Figure 7 shows the local Pareto front for solutions to SDEL_Ω. As a comparison, locally optimal solutions x^*, corresponding to the deterministic plan ($c \equiv 0$), are evaluated in SDEL_Ω. The results show that there is an

unexpected leakage of radiation due to slower moving leaves, and this in turn implies that the organs receive an additional dose. The SMPEC model is to some extent able to cope with this situation and delivers on average a better treatment plan compared to a deterministic model.

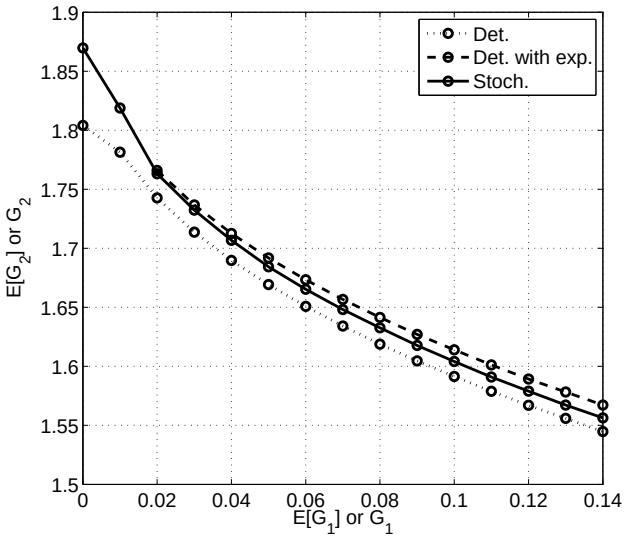


Figure 7: The figure shows three fronts: one is a local Pareto front for solutions x_s^* to the stochastic model SDEL_Ω (Stoch.); one is a local Pareto front for solutions x_d^* to the corresponding deterministic model, i.e., with $c \equiv 0$ (Det.); and one is the deterministic solutions x_d^* evaluated in SDEL_Ω (Det. with exp.).

5 Summary, Conclusions and Future Research

This paper contributes with two rather different applications of robust SMPEC models.

First, we consider the case of robust network design under tolled user equilibrium flows. This is a case where the lower-level problem is a variational inequality over a polyhedral set, and where uncertainty may be present in both the demand and travel cost functions. Our numerical example focuses on the latter, and we experiment with both the traditional average and the CVaR objective. Robustness of stationary solutions follows, since we assume the travel cost to be separable and affine with positive coefficients. In practice, link costs normally are modeled as nonlinear. In order

to establish the robustness of stationary solutions and the convergence of Monte Carlo schemes, Assumption (B) must be enforced (cf. Theorems 2.2 and 2.4), which implies a strong monotonicity assumption on the travel cost; this may be a limiting factor in some applications. A particularly nice feature of the present application is the availability of information from the responses generated in the Monte Carlo scheme, as revealed in Figures 2 and 3. In applications of toll optimization, where equity is an important issue, the CVaR objective provides a very interesting performance measure, as it allows for the optimization of the worst case situation. It would be interesting to further study applications of robust toll setting problems under uncertainty in this setting, in particular in the multiclass, elastic demand setting.

Second, we consider the optimization of a robust treatment plan in intensity modulated radiation therapy. We consider two lower-level problems, each with its unique type of uncertainty: one with a simple linear system of equations, where the position of the tumor is uncertain, and one with a linear program where some of the machine parameters are uncertain. In the first of these examples, both the expected value and the CVaR objective are used to tackle the position uncertainty. The properties of the simple lower-level problem imply that the problem is convex and also that globally optimal solutions are robust. The numerical example shows that the risk of complications can be reduced significantly if the conservative extra margin around the tumor is reduced and tumor coverage is enforced with expected value or CVaR minimization.

The linear program in the lower-level problem represents the problem to find optimal leaf trajectories to block the beam such that a given intensity profile is (almost) attained in minimum time. Compared to the other version with a linear lower-level problem, this formulation is approaching what is called a deliverable treatment. This means that the optimal solution can be used with much less postprocessing, which may otherwise worsen the plan. We assume that the speed of the leaves are uncertain, which implies that there will be unexpected leakages of radiation. Due to the presence of a linear lower-level problem, the resulting SMPEC violates Assumption (B); to establish robustness, we consider a heuristic solution to the lower-level problem. We consider the optimization of a treatment plan as a multiobjective problem with two goals: tumor coverage and minimum risk of complications. It was found that the uncertainty in general reduces the quality of a plan due to the radiation leakage, but that the SMPEC model yields a better treatment plan on average compared to a deterministic model.

We already utilize biological objective functions in our IMRT application, and they are becoming increasingly important in practice. As a future

research task, it would be interesting to consider uncertainty also in the radiobiological parameters. Their inclusion into the SMPEC model should be straightforward.

For the application of the SMPEC model to become more practical, special algorithms need to developed.

Acknowledgements This work was financially partially supported by the Gothenburg Mathematical Modeling Centre (GMMC) at Mathematical Sciences (Chalmers University of Technology and the University of Gothenburg), whose main sponsor is the Swedish Foundation for Strategic Research (SSF). The authors wish to thank Caroline Olsson for valuable help with the biology of radiation therapy and for providing a test case. We finally thank the Associate Editor Franco Giannessi for his very encouraging and constructive remarks on a previous manuscript.

References

[1] Patriksson, M., Wynter, L.: Stochastic mathematical programs with equilibrium constraints. Oper. Res. Lett. **25**(4), 159–167 (1999)

[2] Couchman, P., Kouvaritakis, B., Cannon, M., Prashad, F.: Gaming strategy for electric power with random demand. IEEE Trans. Power Syst. **20**(3), 1283–1292 (2005)

[3] Xu, H.: An MPCC approach for stochastic Stackelberg–Nash–Cournot equilibrium. Optimization **54**(1), 27–57 (2005)

[4] Christiansen, S., Patriksson, M., Wynter, L.: Stochastic bilevel programming in structural optimization. Struct. Multidiscip. Optim. **21**(5), 361–371 (2001)

[5] Evgrafov, A., Patriksson, M., Petersson, J.: Stochastic structural topology optimization: existence of solutions and sensitivity analyses. ZAMM Z. Angew. Math. Mech. **83**(7), 479–492 (2003)

[6] Patriksson, M.: Robust bi-level optimization models in transportation science. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **366**(1872), 1989–2004 (2008)

[7] Cromvik, C., Patriksson, M.: On the robustness of global optima and stationary solutions to stochastic mathematical programs with equilibrium constraints, part 1: Theory. J. Optim. Theory Appl. **144**(3) (2010, to appear)

[8] Patriksson, M.: On the applicability and solution of bilevel optimization models in transportation science: A study on the existence, stability and computation of optimal solutions to stochastic mathematical programs with equilibrium constraints. Transp. Res. **42B**(10), 843–860 (2008)

[9] Evgrafov, A., Patriksson, M.: On the existence of solutions to stochastic mathematical programs with equilibrium constraints. J. Optim. Theory Appl. **121**(1), 65–76 (2004)

[10] Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. **5**(1), 43–62 (1980)

[11] Wardrop, J.G.: Some theoretical aspects of road traffic research. In: Proceedings of the Institute of Civil Engineers, Part II, pp. 325–378 (1952)

[12] Aashtiani, H.Z., Magnanti, T.L.: Equilibria on a congested transportation network. SIAM J. Algebraic Discrete Methods **2**(3), 213–226 (1981)

[13] Patriksson, M.: The Traffic Assignment Problem: Models and Methods. VSP, Zeist (1994)

[14] Birbil, Ş.İ., Gürkan, G., Listeş, O.: Solving stochastic mathematical programs with complementarity constraints using simulation. Math. Oper. Res. **31**(4), 739–760 (2006)

[15] Sheffi, Y.: Urban Transportation Networks. Prentice-Hall, Englewood Cliffs, NJ (1985)

[16] Marcotte, P., Patriksson, M.: Traffic equilibrium. In: C. Barnhart, G. Laporte (eds.) Transportation, *Handbooks in Operations Research and Management Science*, vol. 14, pp. 623–713. North–Holland, Amsterdam (2007)

[17] Sensitivity analysis of separable traffic equilibrium equilibria with application to bilevel optimization in network design. Transp. Res. **41B**(1), 4–31 (2007)

[18] Gill, P.E., Murray, W., Saunders, M.A.: SNOPT: an SQP algorithm for large-scale constrained optimization. SIAM Rev. **47**(1), 99–131 (2005)

[19] Gwinner, J., Raciti, F.: Random equilibrium problems on networks. Math. Comput. Modelling **43**(7–8), 880–891 (2006)

[20] Sumalee, A.: Optimal toll ring design with spatial equity impact constraint: An evolutionary approach. J. East. Asia Soc. Transp. Stud. **5**, 1813–1828 (2003)

[21] Connors, R., Sumalee, A., Watling, D.: Equitable network design. J. East. Asia Soc. Transp. Stud. **6**, 1382–1397 (2005)

[22] Maruyama, T., Sumalee, A.: Efficiency and equity comparison of cordon- and area-based road pricing schemes using a trip-chain equilibrium model. Transp. Res. **41A**(7), 655–671 (2007)

[23] Friesz, T.L., Anandalingam, G., Mehta, N.J., Nam, K., Shah, S.J., Tobin, R.L.: The multiobjective equilibrium network design problem revisited: A simulated annealing approach. European J. Oper. Res. **65**(1), 44–57 (1993)

[24] Migdalas, A.: Bilevel programming in traffic planning: Models, methods and challenge. J. Global Optim. **7**(4), 381–405 (1995)

[25] Larsson, T., Patriksson, M.: Side constrained traffic equilibrium models—traffic management through link tolls. In: P. Marcotte, S. Nguyen (eds.) Equilibrium and Advanced Transportation Modelling, pp. 125–151. Kluwer Academic Publishers, Boston, MA (1998)

[26] Yin, Y.: Robust optimal traffic signal timing. Transp. Res. **42B**(10), 911–924 (2008)

[27] Stavrev, P., Hristov, D., Warkentin, B., Sham, E., Stavreva, N., Fallone, B.G.: Inverse treatment planning by physically constrained minimization of a biological objective function. Med. Phys. **30**(11), 2948–2958 (2003)

[28] Niemierko, A.: Reporting and analyzing dose distributions: A concept of equivalent uniform dose. Med. Phys. **24**(1), 103–110 (1997)

[29] Olafsson, A., Wright, S.J.: Efficient schemes for robust IMRT treatment planning. Phys. Med. Biol. **51**(21), 5621–5642 (2006)

[30] Chu, M., Zinchenko, Y., Henderson, S.G., Sharpe, M.B.: Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. Phys. Med. Biol. **50**(23), 5463–5477 (2005)

[31] Chan, T.C.Y., Bortfeld, T., Tsitsiklis, J.N.: A robust approach to IMRT optimization. Phys. Med. Biol. **51**(10), 2567–2583 (2006)

[32] Baum, C., Alber, M., Birkner, M., Nüsslin, F.: Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities. Radiother. Oncol. **78**(1), 27–35 (2006)

[33] Unkelbach, J., Oelfke, U.: Relating two techniques for handling uncertainties in IMRT optimization. Phys. Med. Biol. **51**(23), 423–427 (2006)

[34] Nakamura, R.A., Monti, C.R., Castilho, L.N., Trevisan, F.A., Valim, A.C., Reinato, J.A.: Prognostic factors for late urinary toxicity grade 2–3 after conformal radiation therapy on patients with prostate cancer. Int. Braz. J. Urol. **33**(5), 652–659 (2007)

[35] Harsolia, A., Vargas, C., Yan, D., Brabbins, D., Lockman, D., Liang, J., Gustafson, G., Vicini, F., Martinez, A., Kestin, L.L.: Predictors for chronic urinary toxicity after the treatment of prostate cancer with adaptive three-dimensional conformal radiotherapy: dose-volume analysis of a phase II dose-escalation study. Int. J. Radiat. Oncol. Biol. Phys. **69**(4), 1100–1109 (2007)

[36] Kåver, G., Lind, B.K., Löf, J., Liander, A., Brahme, A.: Stochastic optimization of intensity modulated radiotherapy to account for uncertainties in patient sensitivity. Phys. Med. Biol. **44**(12), 2955–2969 (1999)

[37] Lian, J., Xing, L.: Incorporating model parameter uncertainty into inverse treatment planning. Phys. Med. Biol. **31**(9), 2711–2720 (2004)

[38] Brahme, A.: Individualizing cancer treatment: biological optimization models in treatment planning and delivery. Int. J. Radiat. Oncol. Biol. Phys. **49**(2), 327–337 (2001)

[39] Webb, S.: Intensity-Modulated Radiation Therapy. Institute of Physics Publishing, London (2001)

[40] Löf, J.: Development of a general framework for optimization of radiation therapy. Ph.D. thesis, Stockholm University, Stockholm (2000)

[41] Carlsson, J.: Utilizing problem structure in optimization of radiation therapy. Ph.D. thesis, Royal Institute of Technology, Stockholm (2008)

[42] Emami, B., Lyman, J., Brown, A., Coia, L., Goitein, M., Munzenrider, J.E., Solin, L.J., Wesson, M.: Tolerance of normal tissue to therapeutic irradiation. Int. J. Radiat. Oncol. Biol. Phys. **21**(1), 109–122 (1991)

[43] Deasy, J., Blanco, A.I., Clark, V.H.: CERR: A computational environment for radiotherapy research. Med. Phys. **30**(5), 979–985 (2003)

[44] Gould, N.I.M., Orban, D., Toint, P.L.: GALAHAD, a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization. ACM Trans. Math. Software **29**(4), 353–372 (2004)

[45] Convery, D.J., Rosebloom, M.E.: The generation of intensity-modulated fields for conformal radiotherapy by dynamic collimation. Phys. Med. Biol. **37**(6), 1359–1374 (1992)

[46] Spirou, S.V., Chui, C.S.: Generation of arbitrary intensity profiles by dynamic jaws or collimators. Med. Phys. **21**(7), 1031–1041 (1994)

[47] Chui, C.S., Spirou, S., LoSasso, T.: Testing of dynamic multileaf collimation. Med. Phys. **23**(5), 635–641 (1996)

[48] Ehrgott, M.: Multicriteria Optimization, second edn. Springer-Verlag, Berlin (2005)

Paper III

Robust multi-objective optimization based on a user perspective

Christoffer Cromvik* Peter Lindroth†

Abstract

Solving practical optimization problems that are sensitive to small changes in the variables or model parameters require special attention regarding the robustness of solutions. We present a new definition of robustness for multi-objective optimization problems. The definition is based on an approximation of the underlying utility function for a single decision maker. We further demonstrate an efficient computational procedure to evaluate robustness. This procedure is applied to two numerical examples: one is an analytic test problem while one is a real-world problem in antenna design. The results show that the robustness varies over the Pareto front and that it can be improved if the decision maker is willing to sacrifice some optimality.

Keywords: Multi-objective optimization, Robustness, Multi-criteria decision making

1 Introduction

Many applications of optimization comprises several more or less conflicting objectives, such as cost/quality, expected return/risk etc. These are to be optimized simultaneously and the aim is to find the most appropriate balance between all the objectives. Mathematically, such a problem is denoted a *multi-objective optimization problem* (MOOP) and is formulated as that to

$$\underset{\mathbf{x}\in X}{\text{minimize}} \ (f_1(\mathbf{x}), \ldots, f_k(\mathbf{x})) \,. \tag{1}$$

*Corresponding author. Department of Mathematical Sciences, Chalmers University of Technology, and Department of Mathematical Sciences, University of Gothenburg, SE-412 96 Gothenburg, Sweden. Tel.: +46-31-7723515, Fax: +46-31-161973, christoffer.cromvik@chalmers.se

†Volvo 3P, Chassis & Vehicle Dynamics, Chassis Strategies & Vehicle Analysis, SE-405 08 Gothenburg, Sweden. Department of Mathematical Sciences, Chalmers University of Technology, and Department of Mathematical Sciences, University of Gothenburg, SE-412 96 Gothenburg, Sweden. Tel.: +46-31-3231063, peter.lindroth@volvo.com

Here, $\mathbf{x} \in \mathbb{R}^n$ denotes a vector of decision variables, $X \subseteq \mathbb{R}^n$ is the feasible decision space, and each $f_i : X \to \mathbb{R},\ i = 1, \ldots, k$, is an objective function to be minimized. Since minimization of a vector in general is not well defined, the notion of optimality for multi-objective problems is somewhat different compared to single-objective problems. Optimality is here based on dominance, and the following definition is used.

Definition 1.1 (Pareto optimality) *A feasible solution $\hat{\mathbf{x}} \in X$ is called* Pareto optimal *if there exists* no *vector $\mathbf{x} \in X$ such that $f_i(\mathbf{x}) \leq f_i(\hat{\mathbf{x}}),\ i = 1, \ldots, k$, with at least one inequality holding strictly. The set of all Pareto optimal solutions is denoted $\mathcal{P} \subseteq X$.*

The possibly most intuitive method for solving a MOOP, i.e., to find $\mathcal{P}$ or at least a good approximation of $\mathcal{P}$, is to solve a sequence of standard optimization problems of the following type

$$\underset{\mathbf{x} \in X}{\text{minimize}} \sum_{i=1}^{k} w_i f_i(\mathbf{x}), \tag{2}$$

where the multiple objectives are transformed to different single objective problems by varying the weight vector $\mathbf{w} \in \{\mathbf{v} \in \mathbb{R}^k \mid \sum_{i=1}^{k} v_i = 1,\ v_i \geq 0\ \forall i\}$. This solution strategy suffers from serious limitations, such that it is only possible to find the subset of $\mathcal{P}$ which is mapped onto the convex part of the Pareto front (Miettinen, 1998), and also that the mapping between $\mathbf{w}$ and the optimal values to (2), i.e., $\mathbb{R}^k \ni \mathbf{w} \mapsto \min_{\mathbf{x} \in X} \sum_{i=1}^{k} w_i f_i(\mathbf{x}) \in \mathbb{R}^k$, is non-linear and strongly depending on the properties of the actual functions involved (Das and Dennis Jr, 1997). To avoid finding *weakly Pareto optimal solutions* (where the strict inequality requirement in Definition 1.1 is dropped), the weights are required to be strictly positive. Despite of its limitations, the weighting strategy is fundamental, and is used as a basis for the definition of robustness presented in this paper.

1.1 Robustness in single- and multi-objective optimization

An optimal solution which is sensitive to perturbations in the data is often not useful in a practical application. A natural approach to deal with this situation is to incorporate the uncertainty into the model. This approach is used in Stochastic Programming (SP) (cf. Kall and Wallace (1994); Birge and Louveaux (1997)) and Robust Optimization (RO) (cf. Ben-Tal and Nemirovski (2002)). In SP, the objective function is typically the expected value over all uncertain parameters, which implies that an optimal solution is good on average. In RO, feasibility is required for all outcomes of the

uncertain parameters, which produces a "conservative" optimal solution. Although most RO theory is restricted to convex problems with an explicit objective function, there are some recent development of RO methods also for non-convex as well as simulation-based problems (cf. Bertsimas et al., 2009)). Das (2000) views robustness as an objective in itself, and sets the goal to generate solutions that optimize both the unperturbed objective value and the expected objective value in a bi-objective optimization fashion.

There are, however, situations where it is not suitable or even possible to remodel the problem, but where there is an interest in assessing the robustness of an optimal solution in a post-process. This opens up the question of how robustness is evaluated. Considering a single-objective problem, we can use the sensitivity of the objective value at an optimal solution as a measure of robustness, but for multi-objective problems this is less straightforward. For such problems we have to quantify the uncertain responses in the objective space; see Figure 1.

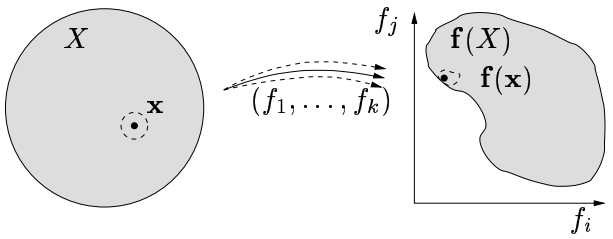


Figure 1: Uncertainties in $\mathbf{x}$ (such as implementation precision) and in $\mathbf{f} = (f_1, \ldots, f_k)$ (e.g. in model parameters) lead to uncertain responses in the objective space.

Among the many papers published on robust optimization, only few concerns multi-objective optimization. One has to distinguish between robust multi-objective optimization for which robustness is one objective and performance is the other (cf. Jin and Sendhoff (2003); Das (2000)), and our interpretation of robust multi-objective optimization where the wish is to find robust solutions to a multi-objective optimization problem. For the latter, Deb and Gupta (2005b,a, 2006) have made a direct extension of SP by using averaged values of the objective functions to define a robust Pareto front.

1.2 Outline

In Section 2, we construct one utility function for each decision maker which measures the objectives. We present a family of utility functions that spans the full range of "hidden objectives", and we also present a few properties of these functions and define two measures of robustness based on them. In Section 3, we discuss the computation of the robustness measures. Depending on the problem in terms of constraints and differentiability, we suggest two approaches to compute approximations of the measures. Section 4 deals with the search for robust solutions. Instead of just assessing the robustness of the Pareto solutions, we state an optimization problem with the goal to find robust, near-optimal solutions. In Section 5 we present two numerical examples. The first uses a known multi-objective test problem and the second considers a real-world problem instance in antenna design. Finally, in Section 6, we summarize the article and suggest some future work.

2 Robustness based on a utility function

To quantify the change in the objective space due to uncertainties in the decision space and in the objective itself, we use the notion of a *hidden objective* in a multi-objective problem. The hidden objective is tailored for each decision maker and captures his/her preferences. The robustness of a solution is then measured by this objective. With this approach, the computation of robustness must be considered as a post-process, since the preferences of the decision maker depend on the Pareto front.

The idea is to present a set of candidate solutions that are robust and constitute a reasonable approximation of the Pareto front. This implies that robustness can be treated as an objective itself, which is natural in a multi-objective setting.

2.1 Hidden objectives in multi-objective optimization

As mentioned previously, a multi-objective problem can often be viewed as a hidden single-objective optimization problem, where hidden means that the objective function is not explicitly known. A decision maker seeks *one* final solution which is optimal to him/her in the sense of balancing the different criteria. The reason for using a multi-objective formulation is to push forward the decisions until more knowledge is revealed about the characteristics and the limitations of the problem at hand. This single-objective optimization problem can be formulated as that to

$$\underset{\mathbf{x}\in X}{\text{minimize}}\ u\left(f_1(\mathbf{x}),\ldots,f_k(\mathbf{x})\right), \tag{3}$$

where $u : \mathbb{R}^k \to \mathbb{R}$ is the hidden single objective. The observation of this formulation is the core of the ideas developed in this paper. Form here on, we refer to the hidden objective as the utility function, and use a convention that a smaller utility value is better than a larger.

Definition 2.1 (rationality) *A utility function* $u : \mathbb{R}^k \to \mathbb{R}$ *is* rational *if for* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{f}(\mathbf{x}) \leq \mathbf{f}(\mathbf{y})$ *implies that* $(u \circ \mathbf{f})(\mathbf{x}) \leq (u \circ \mathbf{f})(\mathbf{y})$. *A decision maker is* rational *if his/her associated utility function is rational.*

Rationality means that if a point $\mathbf{y}$ is dominated by a point $\mathbf{x}$, then $\mathbf{x}$ must be appreciated as at least as good as $\mathbf{y}$.

With the above definition of rationality, the following proposition shows how rational utility functions can be characterized.

Proposition 2.2 *The utility function* u *is rational if and only if* $u(f_1, \ldots, f_k)$ *is monotonically increasing with each* f_i, $i = 1, \ldots, k$.

Proof. If u is monotonically increasing in every argument it holds that $u(\mathbf{f}(\mathbf{x})) \leq u(\mathbf{f}(\mathbf{y}))$ whenever $\mathbf{f}(\mathbf{x}) \leq \mathbf{f}(\mathbf{y})$, i.e., u is rational. Suppose now that u is rational, but not monotonically increasing, i.e., $\exists\ \mathbf{f} \in \mathbb{R}^k$, $j \in \{1, \ldots, k\}$ and $\varepsilon > 0$ such that $u(f_1, \ldots, f_j, \ldots, f_k) > u(f_1, \ldots, f_j + \varepsilon, \ldots, f_k)$. But u is rational and hence since $(f_1, \ldots, f_j, \ldots, f_k) \leq (f_1, \ldots, f_j + \varepsilon, \ldots, f_k)$ it holds that $u(f_1, \ldots, f_j, \ldots, f_k) \leq u(f_1, \ldots, f_j + \varepsilon, \ldots, f_k)$. This is a contradiction, whence u must be monotonically increasing. □

We also make the following assumption on the function values of the Pareto solutions.

Assumption A

The objective values are scaled such that $\mathbf{f}(\mathcal{P}) \subseteq (0, 1]^k$.

If the range of f over X is bounded, it is always possible to scale the objectives such that Assumption A holds true.

2.2 The utility function

We assume that the utility function has the following form:

$$u(\mathbf{f}) := \sum_{i=1}^{k} w_i f_i^{\alpha}, \tag{4}$$

where $\mathbf{w} \in \mathbb{R}^k_+$ are weights and $\alpha \geq 1$ is a parameter related to curvature. We also define a family of utility functions.

Definition 2.3 *A family of* attainable *utility functions* $\mathcal{U}$ *is defined as*

$$\mathcal{U} = \left\{ \sum_{i=1}^{k} w_i f_i^{\alpha} \;\middle|\; w_i > 0,\ i = 1, \ldots, k;\ \alpha \in [1, \infty) \right\}. \tag{5}$$

We associate a utility function to each candidate vector $\bar{\mathbf{x}} \in X$, i.e., to any solution that a decision maker is interested in. If $\bar{\mathbf{x}} \in \mathcal{P}$, then $\mathbf{w}$ and α are chosen such that α is as small as possible and

$$\bar{\mathbf{x}} \in \operatorname{argmin} \{ u \circ \mathbf{f}(\mathbf{x}) \mid \|\nabla_f (u \circ \mathbf{f})(\mathbf{x})\|_1 = 1 \}.$$

In the following, we present a few properties of the family (5) of utility functions. The goals are to show that the family is rational and complete with respect to certain Pareto optimal points in a sense to be defined below. These are points that can be reached using a utility function in the family $\mathcal{U}$, and we will use the notion of *proper Pareto optimality* to identify them. We first define completeness for a general family of utility functions.

Definition 2.4 (completeness) *A family of utility functions* $\mathcal{U}$ *is* complete with respect to a set $\hat{\mathcal{P}} \subseteq \mathcal{P}$ *if for every* $\mathbf{x}^* \in \hat{\mathcal{P}}$ *there exists a* $u \in \mathcal{U}$ *such that*

$$\mathbf{x}^* \in \arg\min_{\mathbf{x} \in X} u\big(f_1(\mathbf{x}), \ldots, f_k(\mathbf{x})\big).$$

That is, in a complete family, for each $\mathbf{x}^* \in \hat{\mathcal{P}} \subseteq \mathcal{P}$ there is at least one utility function that evaluates $\mathbf{x}^*$ as a best one. A good family of utility functions is both rational and complete with respect to a set which is a close approximation to $\mathcal{P}$. We will show that the family (5) is a good one.

Proposition 2.5 *The family of utility functions defined by* (4) *is rational.*

Proof. Since $\mathbf{w} \geq \mathbf{0}^k$ and $\alpha \geq 1$, all $u \in \mathcal{U}$ are monotonically increasing in all their arguments; the result follows then immediately from Prop. 2.2. □

Geoffrion (1968) introduced the notion of *proper* Pareto optimality to exclude some Pareto optimal solutions that are insensible to reasonable decision makers.

Definition 2.6 (proper Pareto optimality) *A feasible solution* $\hat{\mathbf{x}} \in X$ *to* (1) *is called* proper Pareto optimal in the sense of Geoffrion *if it is Pareto optimal in* (1) *and if there exists a number* $M > 0$ *such that for each* $i \in \{1, \ldots, k\}$ *and each* $\mathbf{x} \in X$ *satisfying* $f_i(\mathbf{x}) < f_i(\hat{\mathbf{x}})$, *there exists a* $j \in \{1, \ldots, k\} \backslash \{i\}$ *such that* $f_j(\hat{\mathbf{x}}) < f_j(\mathbf{x})$ *and*

$$\frac{f_i(\hat{\mathbf{x}}) - f_i(\mathbf{x})}{f_j(\mathbf{x}) - f_j(\hat{\mathbf{x}})} \leq M. \tag{6}$$

We denote the set of all proper Pareto vectors in the sense of Geoffrion by $\check{\mathcal{P}}$.

A vector $\mathbf{x}$ is properly Pareto optimal in the sense of Geoffrion if it has finite trade-offs between the objectives. We make a somewhat different definition of proper Pareto optimality based on the family (5) of utility functions.

Definition 2.7 (firmly proper Pareto optimality) *A feasible solution to* (1) *is called* firmly proper Pareto optimal *if it is the minimizer of* (3) *for some utility function* u *in the family* $\mathcal{U}$ *defined in* (5). *We denote the set of all firmly proper Pareto vectors by* $\mathcal{P}'$.

Figure 2 illustrates some firmly proper, proper and non-proper Pareto optimal solutions. The definition of firmly proper Pareto optimal points implies that the family of utility functions is complete with respect to these points. The question now is which points are firmly proper.

The two following propositions show that firmly proper Pareto optimal solutions are indeed Pareto optimal, and that these solutions are also proper in the sense of Geoffrion.

Proposition 2.8 *Under Assumption* A, *each firmly proper Pareto optimal solution is a Pareto optimal solution, i.e.,* $\mathcal{P}' \subseteq \mathcal{P}$.

Proof. Suppose that $\mathbf{x} \in X \backslash \mathcal{P}$. Then, $\exists\, \mathbf{y} \in X$ such that $f_i(\mathbf{y}) \leq f_i(\mathbf{x})$, $i = 1, \ldots, k$, with $f_j(\mathbf{y}) < f_j(\mathbf{x})$ for some index j. It follows that $u(\mathbf{f}(\mathbf{y})) = \sum_{i=1}^{k} w_i f_i(\mathbf{y})^\alpha < \sum_{i=1}^{k} w_i f_i(\mathbf{x})^\alpha = u(\mathbf{f}(\mathbf{x}))$ since $w_j > 0$, $f_j(\mathbf{y}) < f_j(\mathbf{x})$, $\alpha \geq 1$, and $\mathbf{f}(\mathbf{y}) > \mathbf{0}^k$. Thus $\mathbf{x} \notin \mathcal{P}'$ and the proposition follows. □

Proposition 2.9 *Under Assumption* A, *each firmly proper Pareto optimal solution to* (1) *is a proper Pareto optimal solution, i.e.,* $\mathcal{P}' \subseteq \check{\mathcal{P}}$.

Proof. Let $\mathbf{x}^* \in \mathcal{P}'$. Prop. 2.8 implies that $\mathbf{x}^* \in \mathcal{P}$. Suppose that $\mathbf{x}^*$ does not fulfill (6). Then for every $M > 0$, there exists an i and an $\mathbf{x} \in X$ with $f_i(\mathbf{x}) < f_i(\mathbf{x}^*)$ such that $\frac{f_i(\mathbf{x}^*) - f_i(\mathbf{x})}{f_j(\mathbf{x}) - f_j(\mathbf{x}^*)} > M$ for all $j \in \{1, \ldots, k\} \setminus \{i\}$ with $f_j(\mathbf{x}^*) < f_j(\mathbf{x})$.

Let us first consider a problem with two objectives, f_i and f_j. Let (w_i, w_j) and α be the parameters for a utility function with minimum at $\mathbf{x}^*$, and let $f_j(\mathbf{x}^*) > f_j(\mathbf{x}) - \frac{f_i(\mathbf{x}^*)-f_i(\mathbf{x})}{M}$, for every $M > 0$. Then we have that

$$\begin{aligned} u\left(f_i(\mathbf{x}^*), f_j(\mathbf{x}^*)\right) &= w_i f_i(\mathbf{x}^*)^\alpha + w_j f_j(\mathbf{x}^*)^\alpha \\ &\geq \lim_{M\to\infty} w_i f_i(\mathbf{x}^*)^\alpha + w_j \left(f_j(\mathbf{x}) - \tfrac{f_i(\mathbf{x}^*)-f_i(\mathbf{x})}{M}\right)^\alpha \\ &= w_i f_i(\mathbf{x}^*)^\alpha + w_j f_j(\mathbf{x})^\alpha \\ &> w_i f_i(\mathbf{x})^\alpha + w_j f_j(\mathbf{x})^\alpha \\ &= u\left(f_i(\mathbf{x}), f_j(\mathbf{x})\right). \end{aligned}$$

Hence $\mathbf{x}^*$ is not optimal in (3), i.e., $\mathbf{x}^* \notin \mathcal{P}'$. This leads to a contradiction.

Let us consider k objectives. For all $\mathbf{x}$, we can partition the objectives into three sets, $I_1(x) = \{i \mid f_i(\mathbf{x}) < f_i(\mathbf{x}^*)\}$, $I_2(x) = \{j \mid f_j(\mathbf{x}) > f_j(\mathbf{x}^*)\}$ and $I_3(x) = \{e \mid f_e(\mathbf{x}) = f_e(\mathbf{x}^*)\}$. In the inequality chain above, only the indices in I_2 are potentially harmful. But each $j \in I_2$ is above shown to result in a non-strict inequality, and also corresponding to an index resulting in a strict inequality. Therefore, altogether we get $u\big(f_1(\mathbf{x}^*), \ldots, f_k(\mathbf{x}^*)\big) > u\big(f_1(\mathbf{x}), \ldots, f_k(\mathbf{x})\big)$, which is the contradiction sought. □

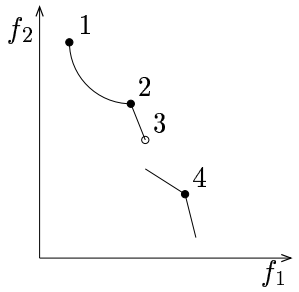


Figure 2: An illustration of the Pareto optimal set for a problem with two objectives. All points except the four marked are proper Pareto optimal points. Points 1 and 2 are not proper, and point 3 is not even Pareto optimal. Point 4 is proper but not firmly proper. Note that point 4 has points arbitrary close on both sides with different values of the trade-offs; this point can therefore be seen as insensible.

We will next identify which points on the Pareto front that are firmly proper. It turns out that for convex multi-objective problems, i.e., with all f_i convex and X convex, it is sufficient with $\alpha = 1$ and $\mathbf{w} \in \mathbb{R}^k_+$ in (4) to make the family (2.3) complete with respect to $\mathcal{P}'$. Since we require that the weights are strictly positive, there may be a few non-proper solutions;

however, almost all Pareto optimal points to convex problems are firmly proper.

In the following proposition and corollary we show that also certain non-convex multi-objective problems have Pareto fronts consisting of only firmly proper Pareto points. The proposition is similar to what is shown in (Li, 1996); however, we assume that the objectives are scaled such that $\mathbf{f}(\mathcal{P}) \subseteq (0,1]^k$. This enables another line of arguments, leading to a significantly shorter proof.

Proposition 2.10 (convexification) *Consider the problem* (1) *under Assumption* A. *Let the Pareto front be parameterized by* $f_k = \phi(f_1, \ldots, f_{k-1})$, *and let* $\mathbf{x}^* \in \mathcal{P}$. *Assume that the local trade-offs on the Pareto front between the pairs of objectives are continuous at* $\mathbf{f}(\mathbf{x}^*)$, *and assume that* ϕ *is twice continuously differentiable. Then, for a sufficiently large* $p \in [0, \infty)$, *the Pareto front of the problem* $\min_{\mathbf{x}\in X}(f_1(\mathbf{x})^p, \ldots, f_k(\mathbf{x})^p)$ *is convex at* $\mathbf{x}^*$.

Proof. Let $\bar{\mathbf{f}} = \{f_1, \ldots, f_{k-1}\}$ and $h(\bar{\mathbf{f}}) = \phi(\bar{\mathbf{f}})^p$. We will show that $\nabla^2_{(\bar{\mathbf{f}}^p)^2} h$ is positive semi-definite at $\mathbf{f}(\mathbf{x}^*)$.

From the chain rule, we have that

$$\frac{\partial h}{\partial (f_j^p)} = \frac{\partial h}{\partial f_j} \frac{1}{p f_j^{p-1}}$$

and that

$$\frac{\partial^2 h}{\partial (f_j^p)^2} = \frac{\partial^2 h}{\partial f_j^2} \frac{1}{p^2 (f_j^{p-1})^2} - \frac{p-1}{p^2 f^{2p-1}} \frac{\partial h}{\partial f_j},$$
$$\frac{\partial^2 h}{\partial f_j \partial f_i} = \frac{\partial^2 h}{\partial f_j \partial f_i} \frac{1}{p^2 f_j^{p-1} f_i^{p-1}}$$

We denote the exponent of a vector to be component wise, and introduce $D = \operatorname{diag}(\bar{\mathbf{f}}^{p-1})^{-1}$ and $E = D^{p-1}$. With these, we have that

$$\nabla_{\bar{\mathbf{f}}^p} h = \frac{1}{p} D \nabla_{\bar{\mathbf{f}}} h, \tag{7a}$$

$$\nabla_{(\bar{\mathbf{f}}^p)^2} h = \frac{1}{p^2} D \nabla^2_{\bar{\mathbf{f}}^2} h D - \frac{p-1}{p} D \operatorname{diag}(\nabla_{\bar{\mathbf{f}}} h) D E. \tag{7b}$$

Now, since

$$\frac{\partial h}{\partial f_j} = p \phi^{p-1} \frac{\partial \phi}{\partial f_j} \quad \text{and} \quad \frac{\partial^2 h}{\partial f_j^2} = p(p-1)\phi^{p-2} (\frac{\partial \phi}{\partial f_j})^2 + p \phi^{p-1} \frac{\partial^2 \phi}{\partial f_j^2},$$

we get

$$\nabla^2_{\bar{\mathbf{f}}^2} h = p(p-1)\phi^{p-2} \nabla_{\bar{\mathbf{f}}} \phi (\nabla_{\bar{\mathbf{f}}} \phi)^{\mathrm{T}} + p \phi^{p-1} \nabla^2_{\bar{\mathbf{f}}^2} \phi.$$

Finally, by inserting the above expression into (7) we get

$$\nabla_{(\bar{\mathbf{f}}^p)^2} h = \frac{p-1}{p}\phi^{p-2} D \nabla_{\bar{\mathbf{f}}}\phi (\nabla_{\bar{\mathbf{f}}}\phi)^{\mathrm{T}} D + \frac{1}{p}\phi^{p-1} D \nabla^2_{\bar{\mathbf{f}}^2}\phi D - \frac{(p-1)^2}{p^2}\phi^{p-1} D \operatorname{diag}(\nabla_{\bar{\mathbf{f}}}\phi) D E.$$

The first term is positive semidefinite, and since $\nabla_{\bar{\mathbf{f}}}\phi < 0$, the last term is also positive semidefinite. When $p \to \infty$, the second term goes to zero faster than the first term, wherefore the result is proved. □

Corollary 2.11 *Let the Pareto front be parameterized by $f_k = \phi(f_1, \ldots, f_{k-1})$, and let $\mathbf{x}^* \in \mathcal{P}$. Assume that the local trade-offs on the Pareto front between all pairs of objectives are continuous at $\mathbf{f}(\mathbf{x}^*)$, and assume that ϕ is twice continuously differentiable. Then each proper Pareto optimal point is a firmly proper Pareto optimal point, and therefore $\check{\mathcal{P}} = \mathcal{P}'$.*

Proof. It is well known (cf. Ehrgott (2005), Thm. 3.11) that all proper Pareto optimal points to convex multi-objective optimization problems can be found using the standard weighting method with non-negative weights. The result follows from Proposition 2.9. □

The corollary implies that all points on sufficiently smooth Pareto fronts are firmly proper, i.e., for problems with such Pareto fronts, our family of utility functions is complete with respect to the whole of $\mathcal{P}$.

To conclude this section, we have shown that the family $\mathcal{U}$ of utility functions is rational and complete with respect to almost all Pareto solutions arising from convex problems and all Pareto fronts that are smooth enough.

2.3 The robustness index

We present two definitions of robustness for a given decision vector: absolute robustness and relative robustness. Both measures are based on the utility function (4), and for both of them, a smaller value means a more robust point.

Definition 2.12 (absolute robustness index) *Let $\bar{\mathbf{x}} \in \mathbb{R}^n$ be the point whose robustness is to be measured, and let $\boldsymbol{\eta} \in \Omega \subseteq \mathbb{R}^m$ be a stochastic variable with mean $\boldsymbol{\eta}_0$. Suppose that $u(\cdot)$ is the utility function associated with $\bar{\mathbf{x}}$. The absolute robustness index of $\bar{\mathbf{x}}$ is defined as*

$$R_A(\bar{\mathbf{x}}) = \mathbf{E}[(u \circ \mathbf{f})(\bar{\mathbf{x}}, \boldsymbol{\eta}) - (u \circ \mathbf{f})(\bar{\mathbf{x}}, \boldsymbol{\eta}_0)].$$

Definition 2.13 (relative robustness index) *Let $\bar{\mathbf{x}} \in \mathbb{R}^n$ be the point whose robustness is to be measured, and let $\boldsymbol{\eta} \in \Omega \subseteq \mathbb{R}^m$ be a stochastic variable with mean $\boldsymbol{\eta}_0$. Suppose that $u(\cdot)$ is the utility function associated with $\bar{\mathbf{x}}$ and that $\mathbf{x}^*(\boldsymbol{\eta}) \in X \cap \arg\min(u \circ \mathbf{f})(\mathbf{x}, \boldsymbol{\eta})$. The relative robustness index of $\bar{\mathbf{x}}$ is defined as*

$$R(\bar{\mathbf{x}}) = \mathbf{E}[(u \circ \mathbf{f})(\bar{\mathbf{x}}, \boldsymbol{\eta}) - (u \circ \mathbf{f})(\mathbf{x}^*(\boldsymbol{\eta}), \boldsymbol{\eta})].$$

Remark 2.14 *Due to Jensen's inequality (cf. Fristedt and Gray (1997), Prop. 12), if $(u \circ \mathbf{f})(\bar{\mathbf{x}}, \cdot)$ is convex, the absolute robustness is non-negative,*

$$R_A(\bar{\mathbf{x}}) \geq (u \circ \mathbf{f})(\bar{\mathbf{x}}, \mathbf{E}[\boldsymbol{\eta}]) - (u \circ \mathbf{f})(\bar{\mathbf{x}}, \boldsymbol{\eta}_0) = 0.$$

In contrast to absolute robustness, relative robustness is not necessarily affected by large changes in the objective space due to different outcomes of $\boldsymbol{\eta}$, since it measures the relative loss to an optimal solution for each $\boldsymbol{\eta}$; see Figure 3.

Which robustness index should be used may be a matter of choice for a decision maker, but practice may motivate the use of one before the other. For example, using relative robustness requires a minimization for each $\boldsymbol{\eta}$ which limits its practical use on some problems. In Section 3, we present procedures for computing approximations of the robustness indices.

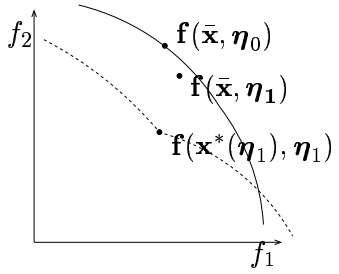


Figure 3: Two Pareto fronts for two realizations of the uncertainty parameter $\boldsymbol{\eta}$. There is a quality loss since the chosen candidate $\bar{\mathbf{x}}$ is not optimal for the outcome $\boldsymbol{\eta}_1$. This quality loss is measured in the relative robustness index.

3 Computation of the utility function and the robustness index

In this section we present practical approaches for computing the utility function and the robustness indices. We start by noting that the computation of robustness is a post-process since it requires a sufficient resolution of the Pareto front. We state this as an assumption:

Assumption B

The Pareto front is computed to a sufficient accuracy and resolution.

The computation of the robustness indices for a specific solution, which we call the candidate, is organized in a series of steps, where the main points are stated in Algorithm 1.

Algorithm 1 Calculate robustness index

Input: Candidate $\bar{\mathbf{x}}$, Pareto front $\mathbf{f}(\mathcal{P})$.

1. Approximate the Pareto front by a quadratic implicit curve around $\bar{\mathbf{x}}$.

2. Compute the utility function u for the candidate such that equations (8) and (9) are fulfilled.

3. Compute R or R_A according to the descriptions in subsections 3.1 and 3.2.

We assume that the Pareto front $\mathbf{f}(\mathcal{P})$ is described by a level set of an implicit function $z(\mathbf{f}(\mathcal{P})) = 0$. By the definition of the utility function u, it is minimized by the candidate $\bar{\mathbf{x}}$. This implies the following two conditions which are also illustrated in Figure 4:

$$\nabla_f u = \gamma \nabla_f z, \tag{8}$$

$$\kappa(u) \geq \kappa(z), \tag{9}$$

where $\gamma \in \mathbb{R}_+$, and $\kappa(\cdot)$ is a measure of curvature. We make a quadratic fit Q of the Pareto front based on the Pareto points within a ball of radius $\tau > 0$. In particular, given a candidate $\bar{\mathbf{x}}$ and the Pareto points in the vicinity, $\mathbf{x}^j$ for $j = 1, \ldots, p$, we solve the following linear least-squares problem

$$\begin{aligned} \underset{c \in \mathbb{R}^m, b \in \mathbb{R}^m}{\text{minimize}} \quad & \sum_{j=1}^{p} \sum_{i=1}^{k} \left(c_i f_i(\mathbf{x}^j)^2 + b_i f_i(\mathbf{x}^j) - 1 \right)^2, \\ \text{subject to} \quad & \sum_{i=1} c_i f_i(\bar{\mathbf{x}})^2 + b_i f_i(\bar{\mathbf{x}}) = 1, \end{aligned}$$

and set

$$Q(\mathbf{f}) = \tfrac{1}{2} \mathbf{f}^{\mathrm{T}} \operatorname{diag}(\mathbf{c})\, \mathbf{f} + \mathbf{b}^{\mathrm{T}} \mathbf{f} - 1.$$

This yields an estimate of the normal and curvature of the front. Since $|\partial Q / \partial f_i| > 0$ and Q is twice continuously differentiable, by the implicit function theorem, there exists a twice continuously differentiable ϕ such that $f_k = \phi(f_1, \ldots, f_{k-1})$. This means that all points on Q are firmly proper according to Proposition 2.10. So even though the Pareto front may not be sufficiently smooth, we are always able to reach all points on the approximate

front, and there always exists an α such that equations (8) and (9) hold. We use normal curvature in equation (9) and we define it, along a vector d, to be

$$\kappa = \frac{d^{\mathrm{T}} H d}{\|d\|^2}, \tag{10}$$

where H is the Jacobian of the normal $N \in \mathbb{R}^k$ to the surface,

$$H = \begin{bmatrix} \frac{\partial N_1}{\partial x_1} & \cdots & \frac{\partial N_1}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial N_k}{\partial x_1} & \cdots & \frac{\partial N_k}{\partial x_k} \end{bmatrix}. \tag{11}$$

We note that in three dimensions, the principal curvatures are the two nonzero eigenvalues to the matrix H (Araujo and Jorge, 2004). Equations (9) and (10) should hold for all directions $d \in \mathbb{R}^k$, although in practice, we only consider a finite set of directions. For the quadratic implicit surface $Q(\mathbf{f})$, equation (11) reduces to (cf. Hughes (2003); Araujo and Jorge (2004))

$$H = \frac{\nabla^2_{ff} Q}{||\nabla_f Q||} - \frac{(\nabla_f Q \nabla_f Q^{\mathrm{T}}) \nabla^2_{ff} Q}{||\nabla_f Q||^3}. \tag{12}$$

A utility function which fulfills equations (8) and (9) can be found by iteratively setting the curvature parameter $\alpha \geq 1$ with corresponding weights $\mathbf{w} > \mathbf{0}$ such that equation (8) is satisfied. Equation (9) will then be satisfied for a sufficiently high value of α.

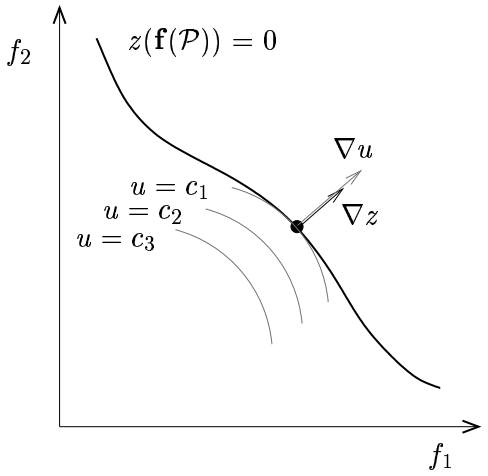


Figure 4: An illustration of the requirements (8) and (9) on the utility function u.

Given a candidate with a corresponding utility function, the next step is to compute R or R_A. This is described in the following subsections. For unconstrained problems with analytic objective functions, we present an approximate closed-form expression for relative robustness. For constrained problems, we show how a Monte-Carlo method can be used. Both methods are used in the numerical experiments in Section 5.

3.1 Unconstrained problem

In addition to assumptions A and B, which we assume holds for $\mathbf{f}(\mathbf{x}, \boldsymbol{\eta}_0)$, we assume the following:

Assumption C

(C1) The functions $f_i(\cdot, \cdot) > 0$, $i = 1, \ldots, k$, are twice continuously differentiable.

(C2) The feasible set is $X = \mathbb{R}^n$.

Under these assumptions, we can formulate a closed-form expression for an approximation of relative robustness. The approximation is based on the second-order Taylor expansion U of the utility function u. With $\hat{u}(\mathbf{x}, \boldsymbol{\eta}) := u(\mathbf{f}(\mathbf{x}, \boldsymbol{\eta}))$, we have

$$\begin{aligned} U(\mathbf{x}, \boldsymbol{\eta}) = \hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0) &+ \nabla_x \hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0)^{\mathrm{T}}(\mathbf{x} - \bar{\mathbf{x}}) + \nabla_\eta \hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0)^{\mathrm{T}}(\boldsymbol{\eta} - \boldsymbol{\eta}_0) \\ &+ (\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{T}} \nabla^2_{x\eta} \hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) + \tfrac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{T}} \nabla^2_{xx} \hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0)(\mathbf{x} - \bar{\mathbf{x}}) \\ &+ \tfrac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)^{\mathrm{T}} \nabla^2_{\eta\eta} \hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0)(\boldsymbol{\eta} - \boldsymbol{\eta}_0), \end{aligned}$$

Since the candidate $\bar{\mathbf{x}}$ is defined to minimize the utility function, the Hessian of u is positive semi-definite. If it is positive definite, and thus non-singular, we get an expression for the optimal solution $\mathbf{x}^*(\boldsymbol{\eta}) \in \arg\min_{\mathbf{x} \in X} U(\mathbf{x}, \boldsymbol{\eta})$ as a (linear) function of the uncertainty parameter $\boldsymbol{\eta}$:

$$\mathbf{x}^*(\boldsymbol{\eta}) = \bar{\mathbf{x}} - \nabla^2_{xx} \hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0)^{-1} \left[\nabla_x \hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0) + \nabla^2_{x\eta} \hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0)^{\mathrm{T}}(\boldsymbol{\eta} - \boldsymbol{\eta}_0) \right].$$

Inserting this into the definition of robustness (2.13) leads to a closed-form expression for the approximate relative robustness index:

$$\begin{aligned} R^U(\bar{\mathbf{x}}) := &\mathbf{E}\left[\hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}) - \hat{u}(\mathbf{x}^*(\boldsymbol{\eta}), \boldsymbol{\eta})\right] = \\ &\mathbf{E}\left[\tfrac{1}{2} \nabla_x \hat{u}^{\mathrm{T}} \nabla^2_{xx} \hat{u}^{-1} \nabla_x \hat{u} + \nabla_x \hat{u}^{\mathrm{T}} \nabla^2_{xx} \hat{u}^{-1} \nabla^2_{\eta x} \hat{u}^{\mathrm{T}}(\boldsymbol{\eta} - \boldsymbol{\eta}_0) \right. \\ &\left. + \tfrac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)^{\mathrm{T}} \nabla^2_{\eta x} \hat{u} \nabla^2_{xx} \hat{u}^{-1} \nabla^2_{\eta x} \hat{u}^{\mathrm{T}}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)\right]. \end{aligned} \tag{13}$$

Introducing Λ as the covariance matrix of $\boldsymbol{\eta}$, and noting that $\nabla_x \hat{u}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0) = 0$ since $\bar{\mathbf{x}}$ is the minimizer, the expression (13) reduces to

$$R^U(\bar{\mathbf{x}}) = \tfrac{1}{2}\mathrm{tr}\left(\Lambda \nabla^2_{\eta x}\hat{u}\nabla^2_{xx}\hat{u}^{-1}\nabla^2_{\eta x}\hat{u}^{\mathrm{T}}\right). \tag{14}$$

This expression only requires the solution of one linear equation with n unknowns and a few matrix-matrix multiplications, and is thus relatively fast to compute.

3.2 Constrained problem

If any of the functions f_i are non-differentiable, if the problem includes constraints or if analytic expressions of the functions f_i are not available, the closed-form expression (14) does not apply. The robustness indices can however be computed using a Monte-Carlo method with randomized sampling. The idea is to draw N i.i.d. samples $\boldsymbol{\eta}_i$, $i = 1, \ldots, N$, of $\boldsymbol{\eta}$ and replace the expected value by the sample mean. We only consider the absolute robustness index, since the relative index would require one minimization computation for each sample. The Monte-Carlo estimate is given by

$$\hat{R}_A(\bar{\mathbf{x}}) = \frac{1}{N}\sum_{i=1}^{N}\left[u\big(\mathbf{f}(\bar{\mathbf{x}}, \boldsymbol{\eta}_i)\big) - u\big(\mathbf{f}(\bar{\mathbf{x}}, \boldsymbol{\eta}_0)\big)\right].$$

4 Search for robust solutions

In Section 3, we assume (Assumption B) that the Pareto front is pre-computed, and the practical computational procedures presented refer to candidates on the front. In reality, however, we may forsake optimality of a solution if robustness can be gained. The idea is to move away from a Pareto optimal solution $\bar{\mathbf{x}}$ on the front and search for robust solutions in its neighborhood. Let $\tau > 0$ be the radius of the ball around $\mathbf{f}(\bar{\mathbf{x}})$ used for the quadratic approximation Q of the front, and let $\varepsilon > 0$. We use the utility function u to define the neighborhood. For absolute robustness we formulate the optimization problem to

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & R_A(\mathbf{x}), \\ \text{subject to} \quad & u\left(\mathbf{f}(\mathbf{x})\right) - u\left(\mathbf{f}(\bar{\mathbf{x}})\right) \leq \varepsilon, \\ & \|\mathbf{f}(\mathbf{x})) - \mathbf{f}(\bar{\mathbf{x}})\| \leq \tau, \\ & \mathbf{x} \in X. \end{aligned} \tag{15}$$

The solution to (15) is the most robust point with at most a decrease of ε in utility compared to $\bar{\mathbf{x}}$ and which is sufficiently close to $\bar{\mathbf{x}}$ in the objective space such that the local approximation of the Pareto front remains valid.

For relative robustness, we have to take into account that inner (non-Pareto) solutions will have a lower utility value for each realization than the optimal solution. Letting $\Delta u(\mathbf{x}) = u(\mathbf{x}, \boldsymbol{\eta}_0) - u(\bar{\mathbf{x}}, \boldsymbol{\eta}_0)$ denote the loss in utility at the unperturbed state, we formulate the optimization problem to

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & R(\mathbf{x}) + \Delta u(\mathbf{x}), \\ \text{subject to} \quad & u\left(\mathbf{f}(\mathbf{x})\right) - u\left(\mathbf{f}(\bar{\mathbf{x}})\right) \leq \varepsilon, \\ & \|\mathbf{f}(\mathbf{x})) - \mathbf{f}(\bar{\mathbf{x}})\| \leq \tau, \\ & \mathbf{x} \in X. \end{aligned}$$

5 Numerical Experiments

The ideas developed in this article have been applied to both the analytical test functions used in the article by Deb and Gupta (2005b) and on functions derived from real-world numerical data used for antenna optimization (Jakobsson et al., 2008a; Stjernman et al., 2009).

The reader should note that the test functions in the first numerical example are designed to illustrate different principal cases when introducing uncertainty in multi-objective problems, and not designed to imitate practical applications. Our intention with this example is to show how our definition of robustness compares to published results.

The theory developed in this article places no theoretical restriction on the number of objectives. However—for illustrational reasons—in the numerical examples we only consider bi-objective problems.

5.1 Analytical functions

Deb and Gupta (2005b) considers uncertainty in the decision space and formulates a program where each objective function is replaced by its respective average computed over a ball around the intended decision variable, i.e.,

$$f_i^{\text{eff}}(\mathbf{x}) = \frac{1}{|\mathcal{B}(\mathbf{x}, \delta)|} \int_{\mathbf{y} \in \mathcal{B}(\mathbf{x}, \delta)} f_i(\mathbf{y}) d\mathbf{y}.$$

The radius of the ball is given by the parameter $\delta > 0$ which is varied in the numerical tests. A larger value of δ smoothens out the functions and makes sharp global optima less attractive. By using this framework, a "robust" Pareto optimal front is always found, but there is no distinction between the points on this front with respect to robustness. Furthermore, there is no continuous grading of robustness of the points that are not in the robust Pareto set. Deb and Gupta also present an alternative robustness model where they enforce robustness of the resulting solutions using a

constraint. The requirement is that the norm of the difference between the (unperturbed) function value and the averaged (or, the worst case) function value must be kept smaller than a certain threshold value. From our point of view, this formulation also suffers from the weakness that it just classifies solutions as robust Pareto optimal or not. It is also possible that large parts of the objective space will not contain any robust solutions if the effect of uncertainty is large. From now on, we will concentrate on Deb and Gupta's first formulation.

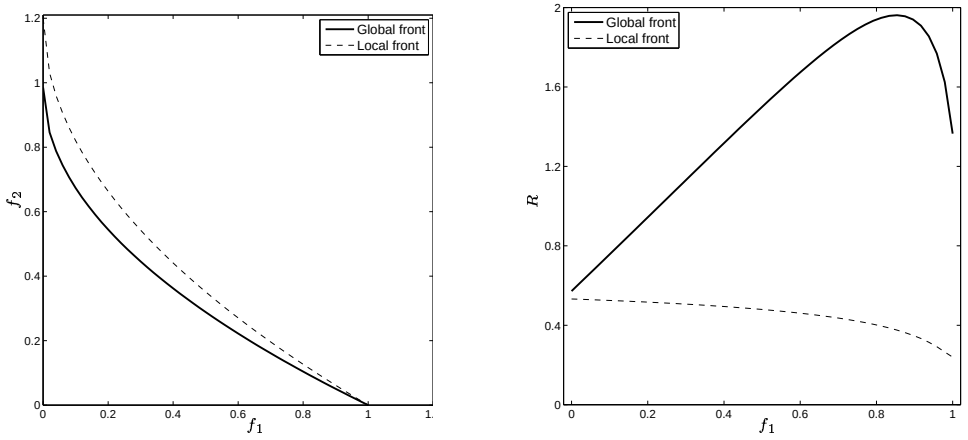
Since we derive robustness for the unperturbed front and Deb and Gupta presents a robust front, possibly consisting of completely different solutions, it is difficult to directly compare the respective results. We will, however, show that they are in line in principle.

We present numerical results for one test problem, DEBGUP3, which is one of 4 bi-objective problems from (Deb and Gupta, 2005b). The problem is to

$$\begin{aligned}
&\underset{\mathbf{x}}{\text{minimize}} \quad (f_1(\mathbf{x}), f_2(\mathbf{x})) = \\
&\qquad \left(x_1, \left(2 - 0.8e^{-\left(\frac{x_2 - 0.35}{0.25}\right)^2} - e^{-\left(\frac{x_2 - 0.85}{0.03}\right)^2} \right) (1 - \sqrt{x_1}) \sum_{i=3}^{5} 50 x_i^2 \right), \\
&\text{subject to} \qquad 0 \le x_i \le 1, \quad i = 1, 2, \qquad \text{(DEBGUP3)} \\
&\qquad\qquad\quad -1 \le x_i \le 1, \quad i = 3, 4, 5.
\end{aligned}$$

The uncertainty appears in the decision space, such that $\mathbf{x}$ is replaced by $\mathbf{x} + \boldsymbol{\eta}$ and $\boldsymbol{\eta}$ is drawn from a uniform distribution, $\boldsymbol{\eta} \in U([-1, 1]^5)$. A close study of the functions reveals that the unperturbed problem with $\boldsymbol{\eta} = \mathbf{0}$ has one local and one global Pareto optimal front, where a local Pareto front consists of points that are locally Pareto optimal. The fronts are shown in Figure 5(a). Figure 5(b) presents the relative robustness index (Def. 2.13) for the corresponding points. We have chosen to ignore the bounds when computing relative robustness which enabled the use of the closed-form expression (14). The implication of ignoring the bounds may be that the value of R is overestimated, i.e., the robustness is underestimated.

Note that the robustness varies both along each single front, and also between the two fronts. The local front is more robust than the global one as is expected from the results in (Deb and Gupta, 2005b). Here, we can distinguish a difference between using the robustness index and using averaged objectives. Depending on the size of the radius δ, the robust front will equal either the local front, the global front, or a combination of these. It is possible to construct problems where the global front equals the robust Pareto front, but having a local Pareto front arbitrarily close and which



(a) Function values.

(b) Robustness indices.

Figure 5: The global and the local Pareto fronts for the problem DEBGUP3. The robustness indices are shown for the corresponding points, parameterized by the values of the first objective.

has much better robustness indices according to our definition. The size of δ highly determines which solutions are presented to a decision maker, whereas the idea in this paper is to partly push forward the decision of how much robustness is desired to the decision maker, and therefore present solutions of different robustness values. The robustness index may also show that robustness may vary along the Pareto front. With more complex objective functions found in real-world applications, we anticipate that there may be more dramatic changes in robustness between close solutions on the Pareto front. In such cases, the decision maker may prefer a solution slightly off his/her ideal (optimal) solution if the robustness properties are better. This situation is presented in the following subsection.

5.2 A real-world example

Designing antennas typically involves a number of conflicting requirements. These may be based on spatial size, so called S-parameters related to electromagnetic properties, functions of the directivity of the antenna, band width, input impedance, or other characteristics of the antenna. In a joint project between the Fraunhofer-Chalmers Centre and the Antenna Research Centre at Ericsson AB a multi-objective optimization approach is taken on the antenna design problem, as described in (Jakobsson et al., 2008a; Stjernman et al., 2009). We have chosen to study this problem using a subset

of the proposed objectives. The decision variables are the positions and geometrical dimensions of the antenna, and the objectives chosen are the maximum return loss ($|S_{11}|$) over the frequency band $[750, 850]$ MHz and the area of the hull of the antenna. An approximate Pareto front is shown in Figure 6, where it is clearly shown that the two objectives are conflicting. The objective functions are expensive to evaluate since they are outcomes of long-running computer simulations. For this reason, a surrogate modeling technique (Jin et al., 2001) is used, where approximate functions are constructed using the function values computed at a number of sample points. Jakobsson et al. (2008a,b) have developed a new technique based on interpolation with rational radial basis functions to handle the sharp function behaviors around the resonance levels.

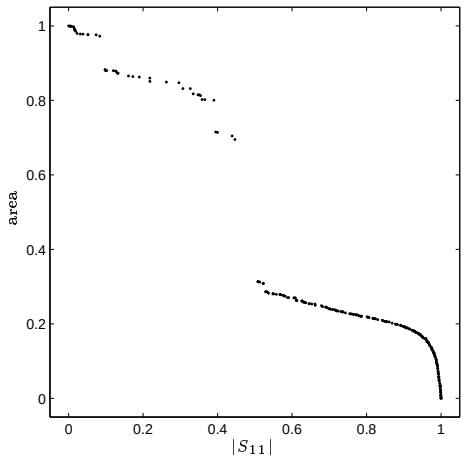


Figure 6: An approximate Pareto front (with the objectives scaled) to the problem found using the NSGA-II algorithm (Deb et al., 2000) with 200 generations and a population size of 48.

The two objectives are interesting for a robustness study. Near resonance, small variations of the decision variables yield large differences in the function values. This is the case for many practical problems where resonance phenomena are part of the problem characteristics. We have noticed that the surrogate models are quite sensitive to the choice of sample points (and this choice is not obvious) and have constructed our numerical study based on this fact.

Originally, the decision space has been sampled at 2000 distinct point chosen using an ad-hoc design–of–experiments strategy, and the surrogate

models (or, response surfaces) have been constructed using the rational RBF technique on the function evaluations at these points. In our experiments, we have randomly selected 500 out of the 2000 points and constructed new response surfaces using only these. The uncertainty characteristics depend on which of the 2000 points that are chosen, reflecting the fact that it is not clear from the start which sample points to choose. Obviously, a *robust* solution is a solution for which the randomness does not have a large effect according to our definition of robustness. In Figure 7, the (absolute)

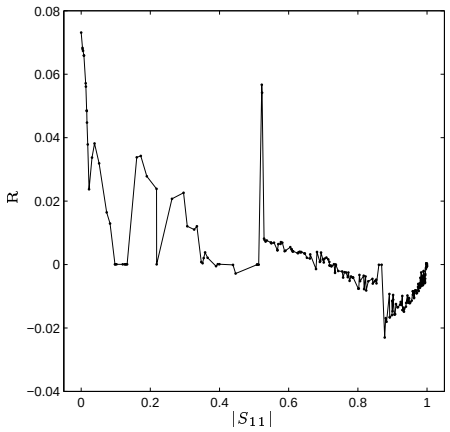
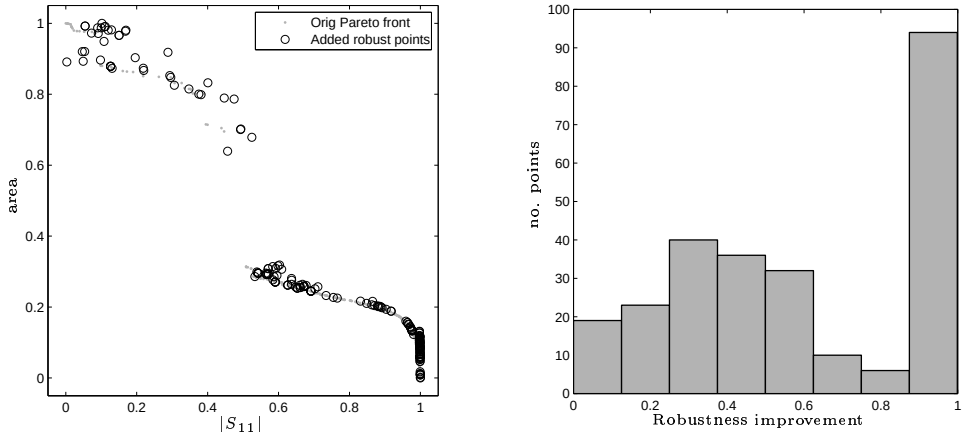


Figure 7: Absolute robustness for the points on the Pareto front, parameterized by the first objective.

robustness index is shown for the points on the Pareto front to the original problem, where the objective functions are the response surfaces constructed using all 2000 data points. The index varies substantially along the front, and for some Pareto points, there are other points on the front that are close in the objective space but with a very different robustness index. This opens up the possibility for a decision maker to choose a point which lies close to his or her ideal point with respect to the function values, but which are much more robust. Doing so will, on average, improve the utility. But since the front is only valid for the unperturbed problem, a decision maker could also be searching for a non-Pareto optimal solution since such a point can be even more robust. In Figure 8, we illustrate such a search. For each (unperturbed) Pareto optimal point, we search for optimal points according to the model (15) with the parameter values $\varepsilon = 0.01$ and $\tau = 0.1$. We use the global optimization algorithm `DIRECT` (Jones, 2009), implemented in `TOMLAB`. We have also implemented a simple local search strategy to complement the

algorithm. In the left figure, the points obtained are shown together with the original (approximate) Pareto front. The right figure shows a histogram of the size of the improvements in robustness when—for each unperturbed point—picking the corresponding robust alternative. One obvious conclusion is that for most Pareto optimal points, there are robust solutions that are close with respect to the value of the utility but with a significantly better robustness index. This fact can be used by a decision maker, who gets an option to choose between robustness and "optimality" for the unperturbed problem.

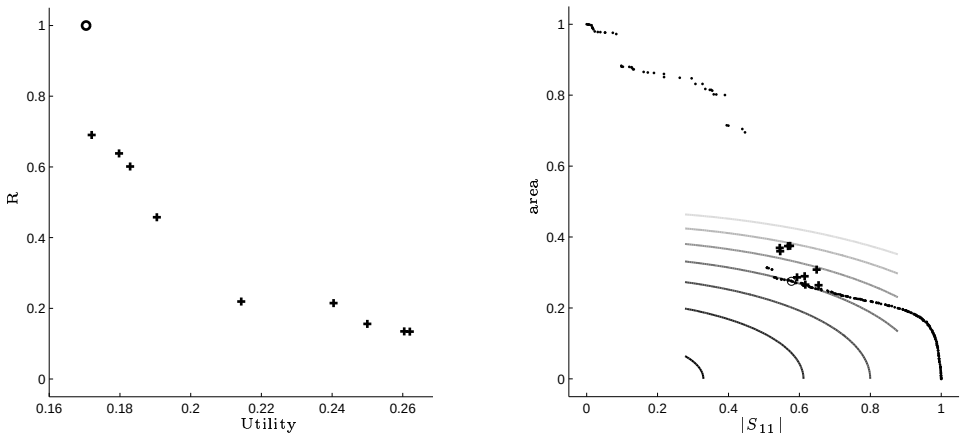


(a) (b)

Figure 8: In a), robust points are added to the (approximate) Pareto front. In b), the relative improvements in the robustness index for the points found are shown.

To further illustrate the framework, we consider the following scenario: Suppose we have presented a Pareto front corresponding to the unperturbed problem to a decision maker, and that he/she has located a candidate solution. Since the problem contains uncertain parameters, the decision maker is also interested in the robustness of this solution. We now solve problem (15) for varying values on the parameter ε. This will produce solutions that are more robust, but with lower utility values. These candidates are then presented to the decision maker, who gets the option to consider how much he or she values robustness considering how much utility is lost. In the spirit of multi-objective optimization, the decision of robustness versus optimality is thus left to the decision maker. Figure 9 shows the results for a specific candidate. In a), the robustness index is shown as a function of the utility for

the alternative solutions. In b), the unperturbed Pareto front is shown along with level curves of the utility function for the specific candidate. The results show that the decision maker can substantially improve the robustness if he or she is willing to sacrifice some utility.



(a) (b)

Figure 9: Figure a) shows the utility function values and robustness index values for the alternative solutions. The robustness index is normalized by the original candidate. Figure b) shows the unperturbed function values for the candidates and level curves of the utility function. In both figures, the ring (o) corresponds to the unperturbed Pareto point originally chosen by the decision maker and the plus signs (+) correspond to the alternative points.

6 Summary and conclusions

We have presented a new definition of robustness for multi-objective problems based on the idea that each decision maker has a hidden single objective function defining which of the Pareto optimal points he/she desires. This hidden objective is characterized with a family of utility functions; we present two robustness indices measuring the relative and absolute expected loss in the utility function due to uncertainties. We have shown that the family of utility functions has certain nice properties such as rationality and completeness. We also presented procedures for computing the robustness indices and applied them to two numerical examples: an analytic test problem from (Deb and Gupta, 2005b), and an in antenna optimization from (Jakobsson et al.,

2008a; Stjernman et al., 2009).

The formulation of robustness by Deb and Gupta (2005b) for multi-objective optimization, which consists of replacing the objectives by their respective expected values, is very natural and direct, and is suitable for many applications. In line with the main idea of multi-objective optimization, our approach has the advantage that the decisions are pushed further into the future when more information about the problem has been revealed. Also, our method gives a continuous measure of robustness and it does that to all points; it does not only say if a point is a robust Pareto optimal point or not.

In future work, our methodology should be applied to more numerical examples, and also to problem instances with more than two objectives. The inclusion of constraints for relative robustness should be developed further. It would also be interesting to apply other types of robustness measures based on utility functions.

Acknowledgments

The first author is partially financially supported by the Gothenburg Mathematical Modeling Centre (GMMC) at Mathematical Sciences (Chalmers University of Technology and the University of Gothenburg), whose main sponsor is the Swedish Foundation for Strategic Research (SSF). The second author is financially supported by Volvo 3P.

The authors wish to thank Michael Patriksson and Ann-Brith Strömberg for their insightful comments and helpful suggestions.

References

B. R. Araujo and J. A. P. Jorge. Curvature dependent polygonization of implicit surfaces. In *Proceedings of the 17th Brazilian Symposium on Computer Graphics and Image Processing*, pages 266–273, Washington DC, USA, 2004. IEEE Computer Society.

A. Ben-Tal and A. Nemirovski. Robust optimization – methodology and applications. *Mathematical Programming*, 92(3):453–480, 2002.

D. Bertsimas, O. Nohadani, and K. M. Teo. Robust nonconvex optimization for simulation-based problems. *Operations Research (to appear)*. URL http://www.optimization-online.org.

D. Bertsimas, O. Nohadani, and K. M. Teo. Nonconvex robust optimization

for problems with constraints. *INFORMS Journal on Computing*, 2009. to appear.

J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research. Springer-Verlag, New York, NY, USA, 1997.

I. Das. Robustness optimization for constrained nonlinear programming problems. *Engineering Optimization*, 32(5):585–618, 2000.

I. Das and J. E. Dennis Jr. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Structural Optimization*, 14:63–69, 1997.

K. Deb and H. Gupta. Handling constraints in robust multi-objective optimization. In *Congress on Evolutionary Computation*, volume 1, pages 25–32, 2005a.

K. Deb and H. Gupta. Searching for robust Pareto-optimal solutions in multi-objective optimization. In *Evolutionary Multi-Criterion Optimization*, pages 150–164, 2005b.

K. Deb and H. Gupta. Introducing robustness in multi-objective optimization. *Evolutionary Computation*, 14(4):463–494, 2006. ISSN 1063-6560.

K. Deb, A. Pratap, Sameer A., and T. Meyarivan. A fast elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6:182–197, 2000.

M. Ehrgott. *Multicriteria Optimization*. Springer-Verlag, Heidelberg, Germany, 2005.

B. Fristedt and L. Gray. *A modern approach to probability theory*. Probability and its Applications. Birkhäuser Boston Inc., Boston, MA, USA, 1997.

A. Geoffrion. Proper efficiency and the theory of vector maximization. *Journal of Mathematical Analysis and Applications*, 22:618–630, 1968.

J. Hughes. Differential geometry of implicit surfaces – a primer. Technical report, Brown University, Computer Science, Providence, USA, 2003.

S. Jakobsson, B. Andersson, and F. Edelvik. Multiobjective optimization applied to design of PIFA antennas. In *Proceedings of the 7th International Conference on Scientific Computing in Electrical Engineering (SCEE 2008)*, Helsinki, Finland, October 2008a.

S. Jakobsson, B. Andersson, and F. Edelvik. Rational radial basis function interpolation with applications to antenna design. Technical report, Fraunhofer Chalmers Research Centre, Göteborg, Sweden, 2008b.

R. Jin, W. Chen, and T. W. Simpson. Comparative studies of metamodeling techniques under multiple modeling criteria. *Journal of Structural Optimization*, 23(1):1–13, 2001.

Y. Jin and B. Sendhoff. Trade-off between optimality and robustness: An evolutionary multiobjective approach, in evolutionary multi-criterion optimization. In *Proceeding of the Second International Conference on Evolutionary Multi-criterion Optimization*, pages 237–251, Algarve, Portugal, 2003. Springer-Verlag.

D. R. Jones. Direct global optimization algorithm. In C.A. Floudas and P.M. Padalos, editors, *Encyclopedia of Optimization*, pages 725–735. Kluwer Academic, 2009.

P. Kall and S. W. Wallace. *Stochastic programming.* Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons Ltd., Chichester, 1994.

D. Li. Convexification of a noninferior frontier. *Journal of Optimization Theory and Applications*, 88(1):177–196, 1996.

K. Miettinen. *Nonlinear Multiobjective Optimization.* Kluwer Academic Publishers, Boston, MA, USA, 1998.

A. Stjernman, A. Derneryd, S. Jakobsson, B. Andersson, and F. Edelvik. Multi-objective optimization of MIMO antenna systems. In *Proceedings of the 3rd European Conference on Antennas and Propagation*, Berlin, Germany, March 2009.

Paper IV

Airbag Folding Based on Origami Mathematics

Christoffer Cromvik* Kenneth Eriksson†

Abstract

A new algorithm for folding three-dimensional airbags is presented. The method is based on Origami mathematics combined with nonlinear optimization.

The airbag is folded to fit into its compartment. Simulating an inflation of the airbag requires an accurate geometric representation of the folded airbag. However, the geometry is often specified in the inflated three-dimensional form, and finding a computer model of the folded airbag is a non-trivial task. The quality of a model of a flat airbag is usually measured by the difference in area between the folded and the inflated airbags.

The method presented here starts by approximating the geometry of the inflated airbag by a quasi-cylindrical polyhedron. Origami mathematics is used to compute a crease pattern for folding the polyhedron flat. The crease pattern is computed with the intention of being fairly simple and to resemble the actual creases on the real airbag.

The computation of the crease pattern is followed by a computation of the folding. This is based on solving an optimization problem in which the optimum is a flat folded model. Finally, the flat airbag is further folded or rolled into its final shape (without using Origami).

The method has been successfully applied to various models of passenger airbags, providing more realistic geometric data for airbag inflation simulations.

1 Introduction

Simulating a crash where the crash test dummy hits an expanding airbag is a challenge to the industry. This situation is called out-of-position (OOP), reflecting that the airbag was not designed for occupants that are sitting too close or for some other reason hit the airbag before it is fully inflated.

*Department of Mathematical Sciences, Chalmers University of Technology and Department of Mathematical Sciences, University of Gothenburg, SE-412 96 Gothenburg, Sweden. E-mail: christoffer.cromvik@chalmers.se

†Department of Engineering Science, University West, SE-461 86 Trollhättan, Sweden. E-mail: kenneth.eriksson@hv.se

The difficulty with an OOP situation compared to an in-position situation is that the inflation of the folded airbag is much more important. It has to be realistically computed, since it affects the impact of the dummy. Attaining a realistic simulation means starting with a correct geometry of the folded airbag and simulating the inflation with correct gas dynamics. Several commercial software packages exist that can simulate the inflation process of an airbag, e.g., the explicit Finite Element (FE) code `LS-DYNA` [5].

This work aims at developing an algorithm for computing an accurate geometry of the flat folded airbag. Different airbags are folded by different methods and with different numbers and types of foldings. The airbags are often folded by both machines and humans according to a folding scheme. Still, the creases are not entirely deterministically positioned. It is very difficult to control the placement of smaller creases. The folding schemes all assume that the airbag lies flat and stretched in some direction. In this position, different foldings are executed until the dimension of the folded airbag is small enough so that it fits into the airbag compartment. The foldings can be a combination of simple folds, but also roll folds.

Some preprocessors to `LS-DYNA`, e.g., `EASi-FOLDER` [4] and `OASYS-PRIMER` [1] contain software for folding a (nearly) flat FE airbag mesh. They are capable of executing the type of foldings that are normally used in production on flat airbags, e.g., roll-fold, z-fold. However, they are not as accurate when presented with the problem of flattening a three-dimensional shape.

Some airbag models have a simple construction, e.g., the driver model which is made of two circular layers sewn together. It is essentially two-dimensional. Passenger airbags are often more complicated. They are made of several layers sewn together in a three dimensional shape, with no trivial two-dimensional representation. See Figure 1 for an example.

In the present work, the computation of the geometry of the flat folded airbag is organized into two steps. First a crease pattern is computed on a polyhedral approximation of the airbag. Second, a nonlinear optimization problem is formed and solved for the purpose of finding the flat geometry. The accuracy of the computed approximation is measured by comparing its area to the area of the inflated model.

2 Crease Pattern

A crease pattern is first designed for a tetrahedron. We present a series of proofs for different types of polyhedra. The proofs are constructive, and their results can be used to design a crease pattern for our application.

Flat foldability, meaning that the polyhedron can be flattened using a

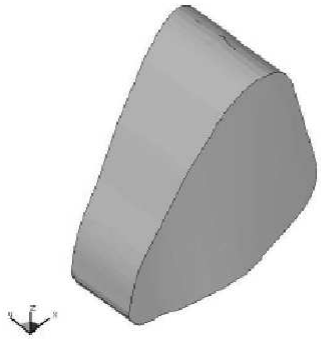


Figure 1: A CAD model of a passenger airbag.

fixed crease pattern, is achieved by cutting along the crease lines, folding the resulting object, and then gluing the cut-up faces back according to the correct connections.

Theorem 2.1. *Any tetrahedron can be folded flat.*

Proof. The proof is organized in a sequence of figures shown in Figure 7, each visualizing the cutting and folding. Consider the tetrahedron with vertices A, B, C, D as in the figure. Cut up the triangle BCD of the tetrahedron, with straight cuts from a point E on the face, to the three vertices B, C, D, respectively, as in the figure.

Then open up the tetrahedron by rotating the triangular patches BDE, BCE, and CDE around the axes BD, BC, and CD, respectively, until these triangles become parts of the three planes through ABD, ABC, and ACD, respectively, as in the figure.

Cut the quadrilateral surface with vertices A, B, E', D along a straight cut from E' to A, and then rotate the resulting triangular faces ABE' and $AE'D$ around the axes AB and AD toward the interior, respectively, until these faces become parts of the two planes ABC and ACD, respectively, as in the figure.

We choose the point E such that the edge BE' after rotation coincides with BE'' and DE' with DE'''. The condition for this is that $\angle ABD + \angle DBE = \angle ABC + \angle CBE$ and $\angle ADB + \angle BDE = \angle ADC + \angle CDE$.

Using this, we may now (partly) restore the surface of the tetrahedron by joining the surfaces ABE'' and $ABE''C$ along the edge BE'', and the surfaces ADE''' and $ADE'''C$ along the edge DE'''.

Finally we rotate the (partly double layered) surface $ADE'''C$ around the axis AC until it coincides with the plane through A,B and C as in

the figure. To conclude the proof of the flat foldability of the tetrahedron we now note that the point E''' after rotation coincides with E''. We may therefore now completely restore the topology of the original tetrahedron by joining the edges AE'' and AE''' (after rotation) and the edges CE'' and CE''' (after rotation). □

Note that the proof is based on cutting and gluing. It does not reveal if there is a continuous deformation to a flat shape.

Remark 2.1. *Concerning the line AE' we remark that the angles $\angle BAE'$ and $\angle DAE'$ satisfy $\angle BAE' + \angle DAE' = \angle BAD$ and $\angle BAC - \angle BAE' = \angle CAD - \angle DAE'$, as in the figure, and are thus independent of the plane BCD. We further note that we may also consider rotating the triangles BDE, BCE and CDE in the opposite direction, again until they become parts of the planes ABD, ABC and ACD, respectively, as in figure. We now choose the point E so that $\angle ABD - \angle DBE = \angle ABC - \angle CBE$ and $\angle ADB - \angle BDE = \angle ADC - \angle CDE$. Continuing from the figure we may then again make a straight cut from E' to A (partly double layered). Again, when we now rotate around the axes AB and AD as before the (rotated) point E' will coincide with E'' and E''' respectively, and we can partly restore the tetrahedron by joining along the edges. Finally, after rotation around AC we may completely restore the topology of the surface of the tetrahedron by joining along the edges. Concerning the crease line from A to E'' we note that again the angles $\angle BAE'$ and $\angle DAE'$ must satisfy the same equations $\angle BAE' + \angle DAE' = \angle BAD$ and $\angle BAC - \angle BAE' = \angle CAD - \angle DAE'$ as before and therefore must be the same as above. We therefore conclude that this crease line is independent of both direction of rotation of the triangles BCE, BDE and CDE, and of the position and orientation of the plane BCD (as long as the angles at A are unchanged).*

We now proceed by cutting the tetrahedron by a plane, see Figure 2. We call the cut-off tetrahedron a prism type polyhedron.

Theorem 2.2. *The prism type polyhedron can be folded flat.*

Proof. Consider a tetrahedron $ABCD$ with the crease pattern from the proof of Theorem 2.1. Cut the tetrahedron with a plane, see Figure 2. In the cut, insert two additional triangular surfaces, such that the two cutoff parts are closed, but not separated. The "smaller" cutoff part is a tetrahedron, and the "bigger" part is a prism type polyhedron. Let the vertices of the smaller tetrahedron be a, b, c, d, where $A = a$, b lies on the edge AB, c on AC and d on AD.

Remark 2.1 shows that the crease line from A to E', see Figure 7, is independent of how the inserted triangular face of the "smaller" tetrahedron

is folded. Let it be folded to the interior of the "smaller" tetrahedron. This means that a crease pattern can be constructed which will coincide with the crease pattern of the original tetrahedron, i.e., the crease line which is constructed by drawing a straight line from a to e' will coincide with the crease line that was created from the line segment from A to E' in the proof of Theorem 2.1.

Now, make an identical copy of the crease pattern on the inserted triangular face belonging to the prism. Folding the original tetrahedron with its inserted triangular faces is possible by the construction of the crease pattern. Let the two polyhedra be separated by moving the tetrahedron in the plane. By the foldability of the tetrahedron, both the smaller tetrahedron and the prism can be folded flat. □

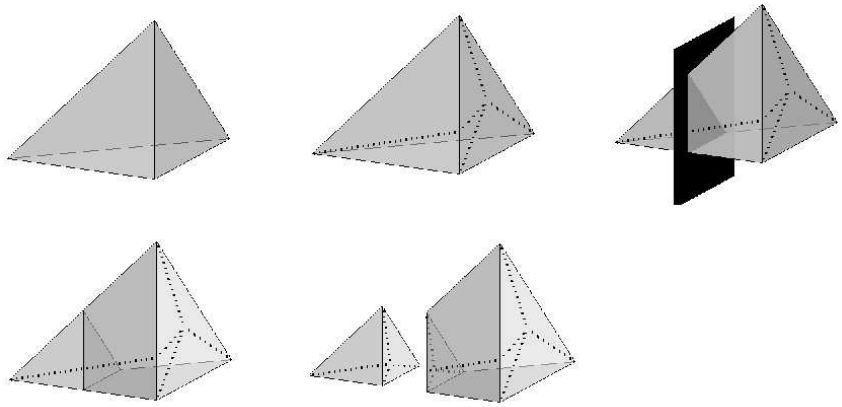


Figure 2: A tetrahedron is cut, and in the cut two additional interior triangular faces are created. Identical crease patterns are created on both interior faces, and the tetrahedron is separated into two parts: a smaller tetrahedron and a prism. The flat foldability of the prism follows from the foldability of the tetrahedron.

Next, we cut the prism type polyhedron by a plane, see Figure 3. We call the cut-off prism a box type polyhedron.

Theorem 2.3. *The box type polyhedron can be folded flat.*

Proof. Let the prism from the cut-off tetrahedron, with its crease pattern, be cut by a plane, see Figure 3. In the cut insert one additional quadrilateral surface which is only connected to the prism by its four vertices. Along the inserted surface put a crease line γ. Its position is only determined by the position of the upper and lower face of the prism. When the prism (with

its cut) and the additional inserted surface are folded, there will be a gap along the sides of the prism, see Figure 2. Let the crease line on the side of the original prism be called ξ. Also, let the point where the crease γ meets ξ unfolded be called p_1, see Figure 2. The gap can be closed by forming two triangles: from a point p, see Figure 2, somewhere along ξ, to the intersection where ξ meets the inserted surface p_2, to B respectively C.

Note that the lengths Cp_1 and Cp_2 are the same, as well as the lengths Bp_1 and Bp_2, and the length Cp is shared by both the gap and the new triangles. Let C_1 and C_2 be positioned according to Figure 2. If the point p is chosen such that $\angle C_1Cp_1 + \angle p_2Cp = \angle C_1CC_2 + \angle C_2Cp$, then the new triangles are an identical match to the gap. By Theorem 2.2, the prism is foldable, so the full construction is foldable, and since the cut does not influence its foldability, and its gap is filled, therefore the box type polyhedron is flat foldable. □

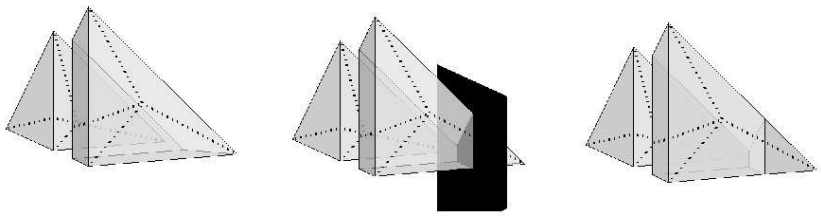


Figure 3: The prism from Figure 2 is cut, and in the cut, an additional interior quadrilateral surface is created. The flat foldability of the box type polyhedron follows from the foldability of the prism and the tetrahedron.

In the proof of Theorem 2.3, a prism was cut off the polyhedron. The process of cutting off a prism can be repeated to create other types of polyhedra.

Definition 2.1. *A quasi-cylindrical polyhedron is a closed cut-off cylinder with a polygonal cross-section.*

Theorem 2.4. *Convex quasi-cylindrical polyhedra are flat foldable.*

Proof. This follows by the proof of Theorem 2.3. In each step, cut off a prism from the polyhedron, until the result forms the given shape. □

Airbags are usually quasi-cylindrical. There are cases, e.g. non-convex polyhedra, for which the technique for generating a crease pattern does not

work. These situations might be avoided by slicing the polyhedron, and computing a crease pattern for each part.

Theorem 2.4 provides an algorithm for designing a crease pattern. Given a quasi-cylindrical polyhedron, we can extend it gradually using prisms until it reaches the shape of a tetrahedron. In each step, we apply the theory for flat foldability, creating a working crease pattern.

3 Folding

For airbags, there are various alternatives for simulating the folding process. This is specially due to the fact that the problem is artificial in the sense that the folding need not be realistic, e.g., there is no need to introduce the concept of time. The objective is to create a flat geometry which is physically correct, not to fold it in a realistic way.

Our algorithm for folding the polyhedron is based on solving an optimization problem. A program is formulated such that the optimal solution represents a flat geometry. The target function, to be minimized, is a sum of rotational spring potentials, one spring over each crease. The minimal value of a spring potential is found when a fold is completed. The constraints are formulated in order to conserve a physically correct representation of the polyhedron, which means conserving the area and avoiding any self-intersections of the faces of the polyhedron.

The crease pattern over a polyhedron induces a subdivision of polygons called patches. In addition, the patches are triangulated, and the interior of the polyhedron is meshed with tetrahedra. Let the nodes of the mesh be $\{x^i\}_{i=1}^{n}$, and let the indices of the surface nodes be I_S. Let the tetrahedra be $\{K_i\}_{i=1}^{n_K}$ and set $I_K = \{1, \ldots, n_K\}$. Let the four indices of the nodes of tetrahedron k be $V_k(i)$, $i = 1, \ldots, 4$. The edges of the triangular faces are denoted $\{E_i\}_{i=1}^{n_E}$, and the indices of the two nodes of edge e are $W_e(i)$, $i = 1, 2$.

Denote the creases $\{C_i\}_{i=1}^{n_C}$. The spring potential over each crease C_i is computed using the scalar product of the normals, n_i^1,n_i^2, of the two neighboring patches. The normals point outward from the polyhedron, and the scalar product is 1 when the two patches are parallel, and -1 when the fold is completed.

The folding process of a polyhedron with n nodes (surface and interior mesh nodes) is formulated as the following nonlinear program with $f : \mathbf{R}^{3n} \to \mathbf{R}$,

$$\underset{x}{\text{minimize}} \quad f(x)$$

$$\begin{aligned} f(x) &= f_1(x) + f_2(x) + f_3(x) \\ &= k_m \sum_{k=1}^{n_K} \left(\sum_{i=1}^{4} \sum_{j=i+1}^{4} \|x^{V_k(i)} - x^{V_k(j)}\| - d_{V_k(i),V_k(j)} \right)^2 \\ &+ \sum_{i=1}^{n_C} n_i^1 \cdot n_i^2 + k_p \sum_{i=1}^{n_E} \left(\|x^{W_i(1)} - x^{W_i(2)}\| - l_{W_i} \right)^2, \end{aligned}$$

subject to

$$\begin{aligned} \text{vol}(K_i) &\geq \varepsilon_1, \qquad i = 1, \ldots, n_K, \\ \text{dist}(x^i, K_j) &\geq \varepsilon_2, \qquad i \in I_S,\ j \in I_K \setminus p_i, \end{aligned}$$

where d_{ij} is the original distance between node x^i and x^j, l_i is the original length of edge i and k_m, k_p are penalty parameters. The first constraint function is $\text{vol}(K_i)$ which is the signed volume of the tetrahedron K_i. The second constraint is $\text{dist}(x^i, K_j)$, which is the distance from a surface node x^i to a tetrahedron K_j, and p_i are the tetrahedron indices connected to node x^i. Finally, ε_1 and ε_2 are small positive constants.

The target function f is composed of three parts. f_1 is a penalty function which strives to keep the tetrahedral mesh uniform. f_2 is the virtual spring potential which drives the folding. f_3 is a penalty function which keeps the edges of the triangles stiff. This is used to maintain the shape and surface area of the patches.

4 Numerical Example

In section 2, a theory for computing a crease pattern was discussed. To demonstrate its practical use, and also to demonstrate the folding algorithm, a numerical experiment is presented. From a CAD-drawing, an airbag shaped polyhedron was constructed. The surface area of the approximation differs about 0.5% to the original area. An in-house optimization solver was used to solve the optimization problem in section 3. It is a Fortran 90 implementation of a low-storage Quasi-Newton SQP method [6, 3, 2], that can handle a few thousand variables and constraints.

The crease pattern was generated by slicing off two upper "bumps", see Figure 5, from the airbag approximation. The crease pattern for these parts were computed separately from the rest of the polyhedron, and the complete crease pattern was formed by joining the parts.

The polyhedron approximation with its crease pattern was meshed using `TetGen` [7]. The visual result (solution) from the optimization progress is shown in Figure 6 for different iteration snapshots.

It was found that the surface area of the flat folded polyhedron was within 0.5% of the surface area of the unfolded polyhedron.

Acknowledgment

The authors would like to thank Prof S. Larsson, Dr B. Pipkorn and K. Mroz for valuable advice.

References

[1] Arup, *Oasys-primer*, `http://www.arup.com`.

[2] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Massachusetts, 2003.

[3] R. Byrd, R. Schnabel, and J. Nocedal, *Representations of quasi-Newton matrices and their use in limited memory methods*, Mathematical Programming **63** (1994), 129–156.

[4] ESI-group, *Easi-folder*, `http://www.esi-group.com`.

[5] Livermore Software Technology Corp., *Ls-dyna*, `http://www.lstc.com`.

[6] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.

[7] Hang S., *TetGen a quality tetrahedral mesh generator and three dimensional Delaunay triangulator version 1.3 user's manual*, Technical report, Weierstrass-Institut für Angewandte Analysis und Stochastik, Berlin, 2004.

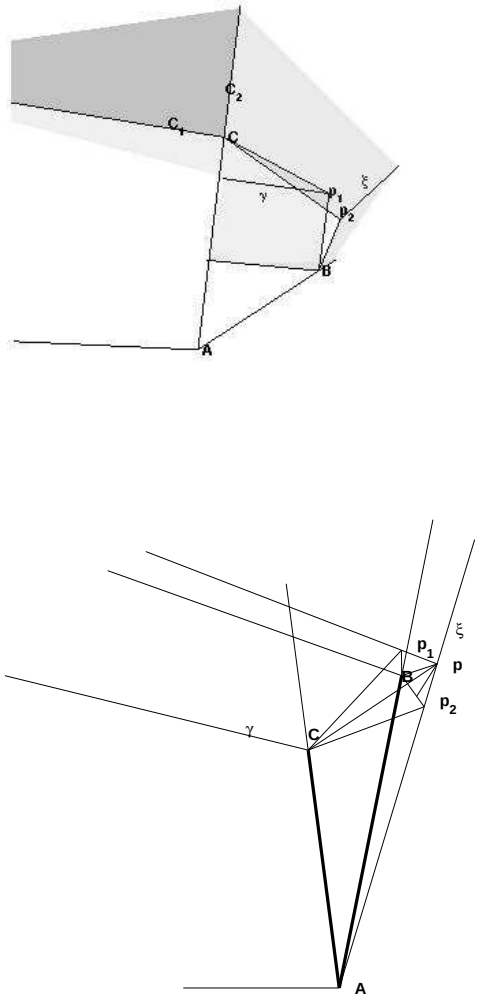


Figure 4: The upper figure shows the gap around the inserted additional surface from the cut. The lower figure shows the same object from above.

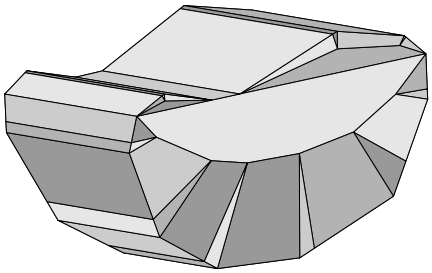


Figure 5: Polyhedral approximation of an airbag model together with a computed crease pattern.

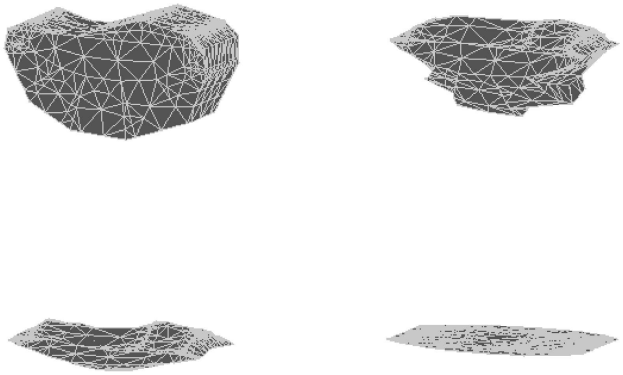


Figure 6: The figures show iteration snapshots from the folding of the polyhedron approximation from Figure 5. The upper left shows the unfolded polyhedron, the upper right: 40 iterations, the lower left: 60 iterations, and the lower right: 200 iterations.

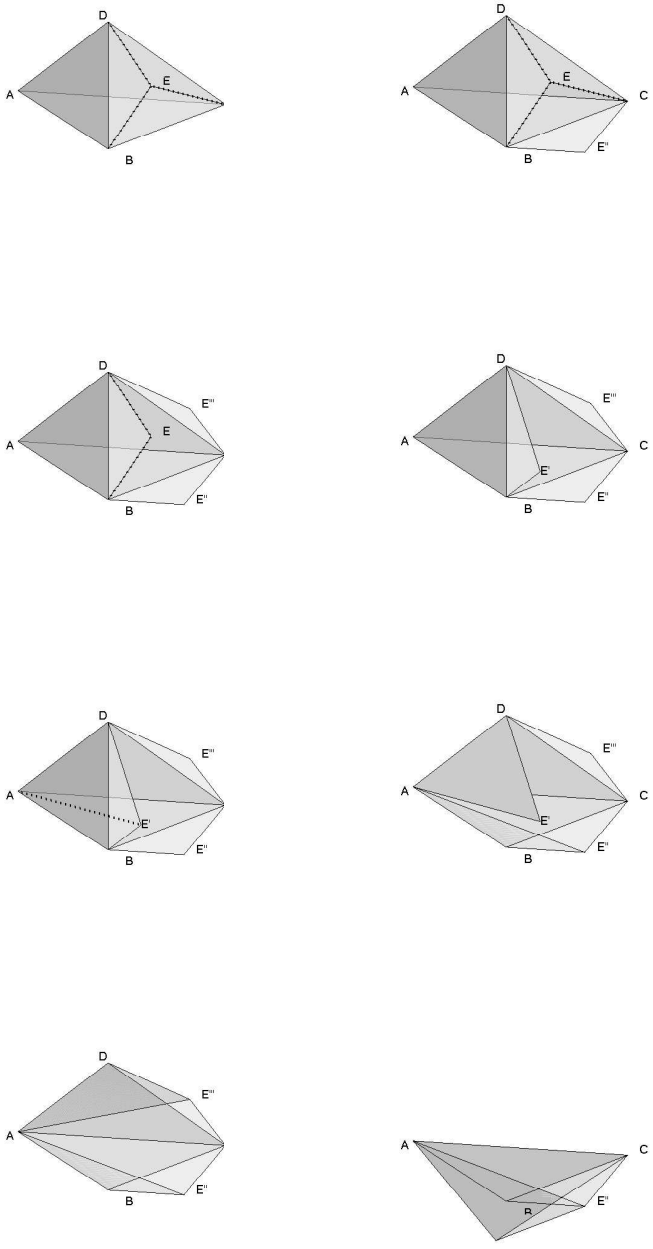


Figure 7: Supporting figure for the proof of Theorem 2.1. The proof follows the figures from left to right beginning at the top.