Fall 2015

Continuous and discrete optimization techniques for some problems in industrial engineering and materials design

Yana Morenko
University of Iowa

CONTINUOUS AND DISCRETE OPTIMIZATION TECHNIQUES FOR SOME PROBLEMS IN INDUSTRIAL ENGINEERING AND MATERIALS DESIGN

by

Yana Morenko

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Industrial Engineering
in the Graduate College of
The University of Iowa

December 2015

Thesis Supervisors: Associate Professor Pavlo Krokhmal
Associate Professor Olesya Zhupanska

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Yana Morenko

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Industrial Engineering at the December 2015 graduation.

Thesis Committee: _____
Pavlo Krokhmal, Thesis Supervisor

_____
Olesya Zhupanska, Thesis Supervisor

_____
Yong Chen

_____
Amaury Lendasse

_____
Albert Ratner

_____
Shaoping Xiao

ACKNOWLEDGEMENTS

I owe my deepest gratitude to my adviser Prof. Krokhmal for his guidance, patience and full support. Additionally, I would like to thank my second adviser, Prof. Zhupanska, for her immense knowledge. I could not imagine having better advisers for my Ph.D. journey. My sincere thank you also goes to the staff of the Department of Mechanical and Industrial Engineering for their help with all organizational questions and warm atmosphere every time I stopped by. Last but not the least, I would like to thank Christopher Kassl for providing datasets necessary for one of the chapters and for supporting me throughout writing this thesis.

ABSTRACT

This work comprises several projects that involve optimization of physical systems. By a *physical system* we understand an object or a process that is governed by physical, mechanical, chemical, biological, etc., laws. Such objects and the related optimization problems are relatively rarely considered in operations research literature, where the traditional subjects of optimization methods are represented by *schedules*, *assignments* and *allocations*, *sequences*, and queues. The corresponding operations research and management sciences models result in optimization problems of relatively simple structure (for example, linear or quadratic optimization models), but whose difficulty comes from very large number (from hundreds to millions) of optimization variables and constraints. In contrast, in many optimization problems that arise in mechanical engineering, chemical engineering, biomedical engineering, the number of variables or constraints in relatively small (typically, in the range of dozens), but the objective function and constraints have very complex, nonlinear and nonconvex analytical form. In many problems, the analytical expressions for objective function and constraints may not be available, or are obtained as solutions of governing equations (e.g., PDE-onstrained optimization problems), or as results of external simulation runs (black-box optimization). In this dissertation we consider problems of classification of biomedical data, construction of optimal bounds on elastic tensor of composite materials, multiobjective (multi-property) optimization via connection to stochastic orderings, and black-box combinatorial optimization of crystal structures of organic molecules.

PUBLIC ABSTRACT

This work comprises four projects that involve optimization of physical systems, which are governed by physical, mechanical, chemical, biological, etc., laws. The first project is focused on efficient solving of special data classification problems, and the developed methodology was applied to several biomedical data sets in order to improve prediction whether a patient or test subject has a certain type of disease (e.g., diabetes).

The second project was concerned with determining and optimizing the ranges of material properties of composite materials. Composite materials typically consist of two or more constituents, which are combined in such a way so as to produce a material whose properties are superior to the properties of the individual constituent materials. The proposed approach was illustrated on nano-composites, or composite materials containing carbon fiber nanotube inclusions.

Third part of this research is focused on multiobjective optimization, that can be used, for example, in market portfolio management.

The final part discusses the possibility of using heuristic algorithms for crystal structure determination from X-ray diffraction data.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ALGORITHMS

Algorithm

CHAPTER 1
INTRODUCTION

In this work we consider several projects, which are relatively disconnected from each other, but have a common theme of dealing with optimization of *physical systems*.

By a *physical system* we understand an object or a process that is governed by physical, mechanical, chemical, biological, etc., laws. Such objects and the related optimization problems are relatively rarely considered in Operations Research literature, where the traditional subjects of optimization methods are represented by *schedules*, *assignments* and *allocations*, *sequences*, and queues. The corresponding Operations Research and Management Sciences problems, such as optimal assignment or allocation of resources, optimal scheduling, and others, result in optimization problems of relatively simple structure (for example, linear or quadratic optimization models), but whose difficulty comes from very large number (from hundreds to millions) of optimization variables and constraints. In contrast, in many optimization problems that arise in mechanical engineering, chemical engineering, biomedical engineering, the number of variables or constraints is relatively small (typically, in the range of dozens), but the objective function and constraints have very complex, nonlinear, and nonconvex analytical form. In many problems, the analytical expressions for objective function and constraints may not be available, or are obtained as solutions of governing equations (e.g., PDE-onstrained optimization problems), or as results of external simulation runs (black-box optimization). This dissertation is concerned with problems of the latter kind. Below we describe the projects that comprise the presented work.

In Chapter 2, we consider a new model of linear separation, which represents one of the popular methods used in data analysis and machine learning. In linear separation problems, the goal is to partition points of a given data set into two classes by a linear surface, or hyperplane. The proposed in Chapter 2 p-norm discrimination model uses a stochastic optimization analogy to treat points that are "misclassified" by a given hyperplane as "outliers" that are assigned special "emphasis" or "weight" and therefore are to be avoided. The amount of "emphasis" can be adjusted by means of the parameter p (the order of the p-norm), such that when $p = 1$, the correctly classified and missclassified points have equal "importance", whereas $p = \infty$ places the largest possible penalty on misclassified points. The proposed approach was tested on three popular datasets that represent various biomedical data. It was shown that linear discrimination models with higher values of parameter $p > 1$ allow for better accuracy of classification comparing to a traditional linear $p = 1$ model. As a byproduct of this project, we also obtained a reformulation of multidimensional p-order cone as an intersection of three-dimensional second-order cones, which is the most compact of such representations currently available in literature.

In Chapter 3, we consider the problem of deriving the tightest possible bounds for material properties of composite materials with randomly distributed inclusions. Composite materials are "hybrid" materials that consist of two or more *phases*, or "ingredients", that act together to overcome each other's weaknesses and form a *composite material* whose properties are superior to those of its constituents. Traditionally, one of the constituent materials is called *matrix*, and other materials represent *inclusions* that are embedded into

the matrix. Composite materials are used in many areas of science and technology, from civil engineering and aerospace industry to dental implants. In development of new composite materials, of great importance is the ability to derive or estimate the properties of the resulting composite from the properties of its constituents. However, the predicted properties of thusly "formulated" composite material may not be achievable with the current manufacturing technology. One example of such a situation is represented by nanocomposites that are based on single-wall carbon nanotubes (CNTs). At the micro-level, CNTs are much stronger than steel due to perfect alignment of nanotubes. At the macrolevel, current manufacturing processes are unable to guarantee a sufficient degree of alignment, which results in a drastic degradation of properties of manufactured samples of CNT buckypaper. Therefore, in this dissertation, we are concerned with construction of the lower and upper bounds on the material properties (i.e., the components of the elastic tensor) of the composite material. Such bounds would represent "the limits of the possible" that can in principle be achieved with the given material phases, regardless of the capabilities or limitations of manufacturing technology. For example, it would be possible to predict whether an improvement of composite's properties could be achieved with an improvement of manufacturing technology, or the selected constituents are unable to yield a composite of required properties regardless of the manufacturing process.

In this chapter, we introduced an optimization approach to determining the lower and upper bounds on the effective overall elastic moduli of a two-phase composite materials that have random distribution/orientation of inclusions. The corresponding optimization model is presented in the form of a nonlinear semidefinite programming (SDP)

problem, where the optimization variable is a tensor of elastic properties. This formulation ensures that the resulting elastic tensors are symmetric and positive semidefinite, which is not always the case in many similar studies in literature. By exploring and exploiting the properties of the objective function and the feasible region of of this nonlinear SDP problem, we were able to reduce the solution process to finding an extremum of a univariate function. In addition, using the obtained optimal solutions we evaluated the quality of well-known Hashin-Strickman-Walpole (HSW) bounds and provided conditions under which these bounds lead to close-to-optimal results, and when they produce bounds that are not guaranteed to correspond to positive semidefinite tensors. Our methods yield the tightest and most consistent bounds existing in the literature.

In Chapter 4 we present a new approach to multiobjective optimization that is based on an analogy with the concepts of stochastic orderings and stochastic dominance relations. This project was originally motivated by development of *multifunctional* materials and structures, capable of performing multiple functions (e.g., structures that can carry load and act like a battery by storing energy) or adapting their performance in response to changes in the operating environment. In the context of development of multifunctional materials, question arises regarding when a material can be considered as truly multifunctional, i.e., at which point do the properties of an engineered material or structure improve significantly enough comparing to the baseline design, so that the resulting design can be considered as multifunctional? In mechanical engineering, improvement of a given property often leads to deterioration of another property. In this regard, a mathematically rigorous methodol-

ogy is required in order to ascertain that an "overall" improvement of a material or design has been achieved, with an improvement in "priority" properties with, perhaps, certain small degradation in other properties. The corresponding optimization models with multiple optimization criteria are known as multi-objective optimization problems. In practice, multiobjective optimization problems are solved by scalarization of the vectorial objective, i.e., by employing a specific transformation that combines several optimization criteria into a single expression. To this end, we proposed a scalarization method for multiobjective optimization that is based on an analogy with stochastic orderings, which are used in decision making under uncertainty to rank multiple randomized outcomes with respect to their "usefulness". Since we were unable to find appropriate material-based data to illustrate our approach, we conducted case studies of the proposed method on multiobjective shortest path problem and portfolio optimization problem with simulated or historical financial data.

Finally, in Chapter 5 we present a problem of optimal reconstruction of crystal structures from X-ray diffraction data. This problem arises in chemistry and chemical engineering, where atoms of chemical elements have to be assigned to "positions" in order to provide the best fit to the observed crystallography data. This problem can naturally be formulated as a black-box combinatorial optimization problem, where an analytic expression for the objective function is not available and is obtained as "external" input. In this project, we proposed several black-box solution algorithms based on popular metaheuristics, such as nearest-neighbor search, simulated annealing, etc.

CHAPTER 2
A P-NORM LINEAR DISCRIMINATION MODEL FOR DATA CLASSIFICATION

2.1 Introduction

Consider two discrete sets $\mathcal{A}, \mathcal{B} \subset \mathbb{R}^n$ containing k and m points, respectively: $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$, $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$. One of the principal tasks arising in machine learning and data mining is that of *discrimination* of such sets, namely, constructing a surface $f(\mathbf{x}) = 0$ such that $f(\mathbf{x}) < 0$ for any $\mathbf{x} \in \mathcal{A}$ and $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{B}$. Of particular interest is the linear separating surface (hyperplane):

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} - \gamma = 0.$$

From the simple fact that any two points $\mathbf{y}_1,\ \mathbf{y}_2 \in \mathbb{R}^n$ satisfying the inequalities $\mathbf{w}^\top \mathbf{y}_1 - \gamma > 0,\ \mathbf{w}^\top \mathbf{y}_2 - \gamma < 0$ for some $\mathbf{w}$ and γ are located on the opposite sides of the hyperplane $\mathbf{w}^\top \mathbf{x} - \gamma = 0$, it follows that the discrete sets $\mathcal{A}$, $\mathcal{B} \subset \mathbb{R}^n$ are considered *linearly separable* if and only if there exist $\mathbf{w} \in \mathbb{R}^n$ such that

$$\mathbf{w}^\top \mathbf{a}_i > \gamma > \mathbf{w}^\top \mathbf{b}_j \quad \text{for all } \ i = 1, \ldots, k,\ j = 1, \ldots, m,$$

with an appropriately chosen γ, or, equivalently,

$$\min_{\mathbf{a}_i \in \mathcal{A}} \ \mathbf{a}_i^\top \mathbf{w} \ > \ \max_{\mathbf{b}_j \in \mathcal{B}} \ \mathbf{b}_j^\top \mathbf{w}. \tag{2.1}$$

Clearly, existence of such a separating hyperplane is not guaranteed (namely, a separating hyperplane exists if the convex hulls of sets $\mathcal{A}$ and $\mathcal{B}$ are disjoint); thus, in general, a separating hyperplane that minimizes some sort of *misclassification error* is desired.

In the next section we introduce a new linear separation model that is based on p-order cone programming, and discuss its key properties. The proposed solution approach, based on a reformulation of p-cone programming problems as second-order cone programming (SOCP) problems when p is rational, is presented in Section 3. Section 4 contains a case study on several popular data sets that illustrates the developed discrimination model.

2.2 The *p*-norm linear separation model: A stochastic optimization analogy

Since definition (2.1) involves strict inequalities, it is not well suited for mathematical programming models of selecting the "best" linear separator. However, the fact that the separating hyperplane can be scaled by any non-negative factor allows one to formulate the following observation:

Proposition 1 (Bennett,Mangasarian [8]). *Discrete sets $\mathcal{A}, \mathcal{B} \subset \mathbb{R}^n$ represented by matrices $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_k)^\top \in \mathbb{R}^{k \times n}$ and $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_m)^\top \in \mathbb{R}^{m \times n}$, respectively, are linearly separable if and only if*

$$\mathbf{A}\mathbf{w} \geq \mathbf{e}\gamma + \mathbf{e}, \quad \mathbf{B}\mathbf{w} \leq \mathbf{e}\gamma - \mathbf{e} \quad \text{for some} \quad \mathbf{w} \in \mathbb{R}^n,\ \gamma \in \mathbb{R}, \tag{2.2}$$

where $\mathbf{e}$ is the vector of ones of an appropriate dimension, $\mathbf{e} = (1, \ldots, 1)^\top$.

Given the linear separability condition (2.2), the (non-negative) vectors

$$\mathbf{x}_{\mathcal{A}} = (-\mathbf{A}\mathbf{w} + \mathbf{e}\gamma + \mathbf{e})_+, \quad \mathbf{x}_{\mathcal{B}} = (\mathbf{B}\mathbf{w} - \mathbf{e}\gamma + \mathbf{e})_+,$$

where $t_+ = \max\{0, t\}$, represent misclassification errors: $\mathbf{x}_{\mathcal{A}}$ and/or $\mathbf{x}_{\mathcal{B}} > \mathbf{0}$ if sets $\mathcal{A}$ and $\mathcal{B}$ are not linearly separable. If one considers that points of sets $\mathcal{A}$ and $\mathcal{B}$ represent

realizations of (discretely distributed) random vectors $\mathbf{a}$, $\mathbf{b} \in \mathbb{R}^n$, respectively, the corresponding elements of vectors $\mathbf{x}_{\mathcal{A}}$, $\mathbf{x}_{\mathcal{B}}$ may be regarded as realizations of random variables $X_{\mathcal{A}}(\mathbf{a}; \mathbf{w}, \gamma) = (-\mathbf{a}^\top \mathbf{w} + \gamma + 1)_+$, $X_{\mathcal{B}}(\mathbf{b}; \mathbf{w}, \gamma) = (\mathbf{b}^\top \mathbf{w} - \gamma + 1)_+$, respectively, that depend parametrically on the decision variables $\mathbf{w}$ and γ. Then, a plausible strategy for selecting $\mathbf{w}$ and γ is one that minimizes, for example, the expected misclassification errors, and which can be formulated as the following stochastic programming problem:

$$\min_{(\mathbf{w},\gamma) \in \mathbb{R}^{n+1}} \Big\{ \delta_1 \mathsf{E}\big[(-\mathbf{a}^\top \mathbf{w} + \gamma + 1)_+\big] + \delta_2 \mathsf{E}\big[(\mathbf{b}^\top \mathbf{w} - \gamma + 1)_+\big] \Big\}, \tag{2.3}$$

where $\delta_{1,2}$ serve as "importance" weights of the misclassification errors for points of sets $\mathcal{A}$ and $\mathcal{B}$, respectively. Further, instead of minimization of expected misclassification error, one may select the parameters $\mathbf{w}$ and γ so as to minimize the risk of misclassification. As it is well known in stochastic optimization and risk analysis, the "risk" associated with random outcome of a decision under uncertainty is often attributed to the "heavy" tails of the corresponding probability distribution. The risk-inducing "heavy" tails of probability distributions, are, in turn, characterized by the distribution's higher moments. Thus, if the misclassifications introduced by a separating hyperplane can be viewed as "random", the misclassification risk may be controlled better if one minimizes not the average, or expected misclassification errors, but their moments of order $p > 1$. This gives rise to the following formulation for linear discrimination of sets $\mathcal{A}$ and $\mathcal{B}$:

$$\min_{(\mathbf{w},\gamma) \in \mathbb{R}^{n+1}} \delta_1 \big\| (-\mathbf{a}^\top \mathbf{w} + \gamma + 1)_+ \big\|_p + \delta_2 \big\| (\mathbf{b}^\top \mathbf{w} - \gamma + 1)_+ \big\|_p, \quad p \in [1, +\infty], \tag{2.4}$$

where $\|\cdot\|_p$ is the usual $\mathcal{L}_p$ norm:

$$\|Y\|_p = \begin{cases} \left(\mathsf{E}|Y|^p\right)^{1/p}, & p \in [1,\infty), \\ \\ \operatorname{ess\,sup}|Y|, & p = \infty. \end{cases}$$

If $\mathbf{a}$ and $\mathbf{b}$ are uniformly distributed with support sets $\mathcal{A}$ and $\mathcal{B}$, respectively:

$$\mathsf{P}(\mathbf{a} = \mathbf{a}_i) = \frac{1}{k}, \quad \mathsf{P}(\mathbf{b} = \mathbf{b}_j) = \frac{1}{m} \quad \text{for all} \quad \mathbf{a}_i \in \mathcal{A}, \;\; \mathbf{b}_j \in \mathcal{B}, \tag{2.5}$$

the p-norm linear discrimination problem takes the form

$$\min_{(\mathbf{w},\gamma)\in\mathbb{R}^{n+1}} \;\; \frac{\delta_1}{k^{1/p}}\big\|(-\mathbf{A}\mathbf{w} + \mathbf{e}\gamma + \mathbf{e})_+\big\|_p + \frac{\delta_2}{m^{1/p}}\big\|(\mathbf{B}\mathbf{w} - \mathbf{e}\gamma + \mathbf{e})_+\big\|_p, \tag{2.6}$$

where $\|\cdot\|_p$ is the vector norm in Euclidean space of appropriate dimension:

$$\|\mathbf{u}\|_p = \begin{cases} (|u_1|^p + \ldots + |u_l|^p)^{1/p}, & p \in [1,\infty), \\ \\ \max\{|u_1|, \ldots, |u_l|\}, & p = \infty, \end{cases}$$

(in the sequel, it shall be clear from the context whether the $\mathcal{L}_p$ or Euclidean p-norm is used). Further, (2.6) can be formulated as a p-order cone programming problem (pOCP)

$$\min \quad \delta_1 k^{-1/p}\,\xi + \delta_2 m^{-1/p}\,\eta \tag{2.7a}$$

$$\text{s.t.} \quad \xi \geq \|\mathbf{y}\|_p, \tag{2.7b}$$

$$\eta \geq \|\mathbf{z}\|_p, \tag{2.7c}$$

$$\mathbf{y} \geq -\mathbf{A}\mathbf{w} + \mathbf{e}\gamma + \mathbf{e}, \tag{2.7d}$$

$$\mathbf{z} \geq \mathbf{B}\mathbf{w} - \mathbf{e}\gamma + \mathbf{e}, \tag{2.7e}$$

$$\mathbf{z},\, \mathbf{y} \geq \mathbf{0}. \tag{2.7f}$$

Note that the special case of $p = 1$ and $\delta_1 = \delta_2$ corresponds to the linear discrimination model of Bennett and Mangasarian [8]. The p-cone programming linear separation model (2.4)–(2.7) shares many key properties with the LP separation model [8], including the guarantee that the optimal solution of (2.7) is non-zero in $\mathbf{w}$ for linearly separable sets.

Proposition 2. *When sets $\mathcal{A}$ and $\mathcal{B}$, represented by matrices $\mathbf{A}$ and $\mathbf{B}$, are linearly separable (i.e., they satisfy* (2.1) *and* (2.2)*), the separating hyperplane $\mathbf{w}^{*\top}\mathbf{x} = \gamma^*$ given by an optimal solution of* (2.4)–(2.7) *satisfies $\mathbf{w}^* \neq \mathbf{0}$.*

Proof. Zero optimal value of (2.7a) immediately implies that at optimality $\mathbf{y}^* = \mathbf{z}^* = \mathbf{0}$, or, equivalently, $-\mathbf{A}\mathbf{w}^* + \mathbf{e}\gamma^* + \mathbf{e} \leq \mathbf{0},\ \mathbf{B}\mathbf{w}^* - \mathbf{e}\gamma^* + \mathbf{e} \leq \mathbf{0}$. If one assumes that $\mathbf{w}^* = \mathbf{0}$, then the above inequalities require that $\gamma^* \leq -1,\ \gamma^* \geq 1$. The contradiction furnishes the desired statement.

Secondly, the p-norm separation model (2.7) can produce a solution with $\mathbf{w} = \mathbf{0}$ only in a rather special case that is identified by Theorem 1 below.

Theorem 1. *Consider the p-order cone programming problem* (2.7)–(2.6)*, where it is assumed without loss of generality that $0 < \delta_1 < \delta_2$. Then, for any $p \in (1, \infty)$ the p-order cone programming problem* (2.7) *has an optimal solution with $\mathbf{w}^* = \mathbf{0}$ if and only if*

$$\frac{\mathbf{e}^\top}{k}\mathbf{A} = \mathbf{v}^\top\mathbf{B}, \quad \textit{where} \quad \mathbf{e}^\top\mathbf{v} = 1,\ \ \mathbf{v} \geq \mathbf{0},\ \ \|\mathbf{v}\|_q \leq \frac{\delta_2}{\delta_1 m^{1/p}}, \tag{2.8a}$$

where q satisfies $\dfrac{1}{p} + \dfrac{1}{q} = 1$. In other words, the arithmetic mean of the points in $\mathcal{A}$ must be equal to some convex combination of points in $\mathcal{B}$. In the case of $\delta_1 = \delta_2$ condition (2.8a) *reduces to*

$$\frac{\mathbf{e}^\top}{k}\mathbf{A} = \frac{\mathbf{e}^\top}{m}\mathbf{B}, \tag{2.8b}$$

i.e., the arithmetic means of the points of sets $\mathcal{A}$ and $\mathcal{B}$ must coincide.

Proof. First, let us consider the case when the p-cone discrimination model (2.7) has an optimal solution with $\mathbf{w}^* = \mathbf{0}$ and demonstrate that (2.8) must then hold. From the formulation (2.6) of problem (2.7) it follows that in the case when $\mathbf{w} = \mathbf{0}$ at optimality, the corresponding optimal value of the objective (2.7a) is determined as

$$\min_{\gamma\in\mathbb{R}} \quad f(\gamma) = \frac{\delta_1}{k^{1/p}}\left(\sum_{i=1}^{k}(1+\gamma)_+^p\right)^{1/p} + \frac{\delta_2}{m^{1/p}}\left(\sum_{j=1}^{m}(1-\gamma)_+^p\right)^{1/p}.$$

Clearly,

$$f(\gamma) = \begin{cases} \delta_1(1+\gamma), & 1 \leq \gamma < \infty, \\ \delta_1 + \delta_2 + \gamma(\delta_1 - \delta_2), & -1 < \gamma < 1, \\ \delta_2(1-\gamma), & -\infty < \gamma \leq -1, \end{cases}$$

whence $\min\limits_{\gamma\in\mathbb{R}} f(\gamma) = f(1) = 2\delta_1$ due to the assumption $0 < \delta_1 < \delta_2$. Next, consider the dual of the p-cone programming problem (2.7):

$$\begin{aligned} \max \quad & \mathbf{e}^\top\mathbf{u} + \mathbf{e}^\top\mathbf{v} \\ \text{s.t.} \quad & -\mathbf{A}^\top\mathbf{u} + \mathbf{B}^\top\mathbf{v} = \mathbf{0}, \\ & \mathbf{e}^\top\mathbf{u} - \mathbf{e}^\top\mathbf{v} = \mathbf{0}, \\ & \mathbf{0} \leq \mathbf{u} \leq -\mathbf{s}, \\ & \mathbf{0} \leq \mathbf{v} \leq -\mathbf{t}, \\ & \|\mathbf{s}\|_q \leq \delta_1 k^{-1/p}, \\ & \|\mathbf{t}\|_q \leq \delta_2 m^{-1/p}, \end{aligned} \tag{2.9}$$

where q is such that $1/p + 1/q = 1$. Note that (2.7) is strictly feasible and bounded from below, since for any $\mathbf{w}_0$, γ_0 and $\varepsilon > 0$ one can select $\mathbf{y}_0 = \varepsilon\mathbf{e} + (-\mathbf{A}\mathbf{w}_0 + \mathbf{e}\gamma_0 + \mathbf{e})_+ > \mathbf{0}$, $\mathbf{z}_0 = \varepsilon\mathbf{e} + (\mathbf{B}\mathbf{w}_0 - \mathbf{e}\gamma_0 + \mathbf{e})_+ > \mathbf{0}$, $\xi_0 = (1+\varepsilon)\|\mathbf{y}_0\|_p > \|\mathbf{y}_0\|_p > 0$, and $\eta_0 = (1+\varepsilon)\|\mathbf{z}_0\|_p >$

$\|\mathbf{z}_0\|_p > 0$ that are feasible to (2.7). Thus, the duality gap for the primal-dual pair of p-order cone programming problems (2.7) and (2.9) is zero [31]. Then, from the first two constraints of (2.9) we have $\mathbf{A}^\top \mathbf{u}^* = \mathbf{B}^\top \mathbf{v}^*$ as well as $\mathbf{e}^\top \mathbf{u}^* = \mathbf{e}^\top \mathbf{v}^*$, which, given that the optimal objective value (2.9) is $2\delta_1$, implies that an optimal $\mathbf{u}^*$ must satisfy

$$\mathbf{e}^\top \mathbf{u}^* = \delta_1. \tag{2.10a}$$

Also, from (2.9) it follows that

$$\|\mathbf{u}^*\|_q \leq \delta_1 k^{-1/p}. \tag{2.10b}$$

Then, it is easy to see that the unique solution of system (2.10) is

$$\mathbf{u}^* = \frac{\delta_1}{k}\mathbf{e} = \left(\frac{\delta_1}{k}, \ldots, \frac{\delta_1}{k}\right)^\top, \tag{2.11}$$

which corresponds to the point where the surface $(u_1^q + \ldots + u_k^q)^{1/q} = \delta_1 k^{-1/p}$ is tangent to the hyperplane $u_1 + \ldots + u_k = \delta_1$ in the positive of $\mathbb{R}^k$.

Similar, an optimal $\mathbf{v}^*$ must satisfy $\mathbf{e}^\top \mathbf{v}^* = \delta_1$ and $\|\mathbf{v}^*\|_q \leq \delta_2 m^{-1/p}$. Note, however, that in the case when $\delta_2/\delta_1 > 1$ such $\mathbf{v}^*$ is not unique. By substituting the obtained characterizations for $\mathbf{u}^*$ and $\mathbf{v}^*$ in the constraint $\mathbf{A}^\top \mathbf{u}^* = \mathbf{B}^\top \mathbf{v}^*$ of the dual (2.9) and dividing by δ_1, we obtain (2.8a). When $\delta_1 = \delta_2$, the optimal $\mathbf{v}^*$ is unique: $\mathbf{v}^* = \frac{\delta_1}{m}\mathbf{e}$, and yields (2.8b).

To prove the statement of the Theorem in the opposite direction, assume that, for instance, (2.8a) holds for certain $\mathbf{u}$ and $\mathbf{v}$. Selecting $\mathbf{u}^* = \frac{\delta_1}{k}\mathbf{e}$, $\mathbf{v}^* = \delta_1 \mathbf{v}$, and $\mathbf{s}^* = -\mathbf{u}^*$, $\mathbf{t}^* = -\mathbf{v}^*$, it is easy to see that $(\mathbf{u}^*, \mathbf{v}^*, \mathbf{s}^*, \mathbf{t}^*)$ represents a feasible solution of the dual problem (2.9) with the dual cost of $2\delta_1$. Similarly, the tuple $(\mathbf{w}^*, \gamma^*, \mathbf{y}^*, \mathbf{z}^*, \xi^*, \eta^*)$, where

$\mathbf{w}^* = \mathbf{0}, \quad \gamma^* = 1, \quad \mathbf{y}^* = (\mathbf{e}\gamma^* + \mathbf{e})_+ = 2\mathbf{e}$, $\mathbf{z}^* = (-\mathbf{e}\gamma^* + \mathbf{e})_+ = \mathbf{0}$, $\xi^* = \|\mathbf{y}^*\|_p = 2k^{1/p}$, $\eta^* = \|\mathbf{z}^*\|_p = 0$, represents a feasible solution of the primal problem (2.7) with the corresponding objective value of $2\delta_1$. Noting the zero duality gap for the constructed pair of feasible solutions of (2.7) and (2.9), and recalling that the primal problem is bounded and strictly feasible, we immediately obtain that this pair of primal-dual solutions is optimal [31]. Hence, from (2.8a) it follows that an optimal solution of (2.7) exists with $\mathbf{w}^* = \mathbf{0}$.

Observe that Theorem 1 implies that in the case of $\delta_1 = \delta_2$ (i.e., when misclassification of points in one set is not favored over that for points of the other set), the p-norm discrimination model (2.7) produces a separating hyperplane with $\mathbf{w} = \mathbf{0}$ only when the "geometric centers" (arithmetic means) of the sets $\mathcal{A}$ and $\mathcal{B}$ coincide. In many situations, this would mean that the convex hulls of the sets $\mathcal{A}$ and $\mathcal{B}$ "overlap" significantly. In practice, this implies that such sets, indeed, cannot be efficiently separated, at least by a hyperplane, thus an occurrence of a $\mathbf{w}^* = \mathbf{0}$ solution in (2.7) should not be regarded as the shortfall of the particular formulation (2.7), but rather the general inapplicability of the linear discrimination method to the specific sets $\mathcal{A}$ and $\mathcal{B}$.

In the case when a "bias" with regard to the importance of misclassification of points of sets $\mathcal{A}$ and $\mathcal{B}$ needs to be introduced by setting $\delta_2 > \delta_1$, occurrence of a $\mathbf{w}^* = \mathbf{0}$ solution in (2.7) does not necessarily imply that sets $\mathcal{A}$ and $\mathcal{B}$ are hardly amenable to linear separation. Indeed, in this case Theorem 1 only claims that the "geometric center" of set $\mathcal{A}$ must coincide with some convex combination of points of set $\mathcal{B}$, i.e., it must coincide with some point inside the convex hull of set $\mathcal{B}$. In this case, linear discrimination can still be a feasible approach, albeit at a cost of significant misclassification errors.

In order to have the stricter condition (2.8b) for the occurrence of $\mathbf{w}^* = \mathbf{0}$ solution in the situation when the preferences for misclassification error are different for sets $\mathcal{A}$ and $\mathcal{B}$, the p-norm linear discrimination model can be extended to the case where misclassifications of points in $\mathcal{A}$ and $\mathcal{B}$ are measured using norms of different orders:

$$\min_{(\mathbf{w},\gamma)\in\mathbb{R}^{n+1}} k^{-1/p_1}\left\|(-\mathbf{A}\mathbf{w}+\mathbf{e}\gamma+\mathbf{e})_+\right\|_{p_1} + m^{-1/p_2}\left\|(\mathbf{B}\mathbf{w}-\mathbf{e}\gamma+\mathbf{e})_+\right\|_{p_2}, \quad p_{1,2}\in(1,\infty). \tag{2.12}$$

Intuitively, a norm of higher order places more "weight" on the outliers; for instance, application of $p = 1$ norm entails minimization of the average misclassification error, in effect regarding all misclassifications as equally important. In contrast, application of the $p = \infty$ norm implies minimization of the largest misclassification errors for the two sets. Thus, by selecting appropriately the orders p_1 and p_2 in (2.12) one may introduce tolerance preferences on misclassifications in sets $\mathcal{A}$ and $\mathcal{B}$. At the same time, it can be shown that the occurrence of $\mathbf{w}^* = \mathbf{0}$ solution in (2.12) would signal the presence of the aforementioned singularity about the sets $\mathcal{A}$ and $\mathcal{B}$. Namely,we have

Theorem 2. *The p-order cone programming problem* (2.12)*, where $p_1, p_2 \in (1,\infty)$, has an optimal solution with $\mathbf{w}^* = \mathbf{0}$ if and only if* (2.8b) *holds.*

In the next section we discuss the details of practical implementation of the p-norm linear discrimination model (2.7).

2.3 A second-order cone programming approach to *p*-order cone programming problems

The p-order cone constraints (2.7b)–(2.7c) are central to practical implementation of the p-norm separation method (2.7) . In the special cases of $p = 1$ or $p = \infty$, p-order cone constraints reduce to linear inequalities; specifically, the $p = 1$ version of model (2.7) has been studied in [8]. In general, the amenability of the 1-norm to implementation via linear constraints has been exploited in a variety approaches and applications, too numerous to cite here. Another prominent special case of is that of $p = 2$, when (2.7b)–(2.7c) represent second-order, or quadratic cones. The second-order cone programming (SOCP) constitutes a well-developed subject of convex optimization, and a number of efficient self-dual "long-step" interior point (IP) SOCP algorithms have been developed in the literature and implemented in software [2, 3, 32, 33, 38]. From the computational standpoint, the "general" case of $p \in (1, 2) \cup (2, \infty)$, when the p-cone is not self-dual, has received relatively little attention in the literature. IP approaches to p-order cone programming have been considered in [16, 34, 39, 43]; an approach based on construction of polyhedral approximations of p-cones and solving the resulting linear programming problems using a cutting-plane technique was proposed in [28].

In this work, we pursue an approach to solving p-cone programming problems that is based on the possibility to represent a p-order cone via a sequence of second-order cones when p is rational [2, 5, 31]. Reformulation of a rational-order p-cone programming problem as a SOCP problem allows for employing the efficient self-dual SOCP methods, but this ability comes at a cost of a large number of second-order cones required for such a

reformulation. In general, a rational-order cone in $\mathbb{R}^{n+1}$ can be represented with $O\big(n\theta_p\big)$ three-dimensional second-order cones, where θ_p is some constant dependent on p. However, such a representation is not unique, and, depending on its particular implementation and the corresponding value of θ_p, the resulting number of second-order cones can vary by $O(n)$. In view of this, in Section 3.2 we introduce a constructive "economical" representation of rational-order p-cones via second-order cones, which facilitates the use of SOCP methods and solvers for tackling p-order cone programming problems.

2.3.1 Representation of rational-order *p*-cones with second-order cones

Without loss of generality, consider a p-cone in the positive orthant of $\mathbb{R}^{n+1}$

$$t \geq (w_1^p + \ldots + w_n^p)^{1/p}, \quad (t, w_1, \ldots, w_n)^\top \geq \mathbf{0}. \tag{2.13}$$

In the case when the parameter p is a positive rational number, $p = r/s$, where $r, s \in \mathbb{N}$, the following "lifted" representation of the p-cone set (2.13) can be constructed in $\mathbb{R}_+^{2n+1}$ [2, 5, 31]:

$$t \geq u_1 + \ldots + u_n, \quad w_j^r \leq u_j^s t^{r-s}, \quad u_j \geq 0, \quad j = 1, \ldots, n. \tag{2.14}$$

Then, each nonlinear constraint in (2.14) can equivalently be replaced by a sequence of inequalities of the form $z^2 \leq xy$, or three-dimensional rotated quadratic cones. Such a representation, as it has been mentioned, is not unique. One way possibility is to rewrite each nonlinear inequality in (2.14) as

$$w^R \leq u^s t^{r-s} w^{R-r}, \tag{2.15}$$

where $R = 2^\rho$, $\rho = \lceil \log_2 r \rceil$, and the subscripts j are suppressed for brevity. Observe that each side of inequality (2.15) contains 2^ρ factors; this allows one to construct a lifted representation for (2.15) via $2^\rho - 1$ three-dimensional rotated quadratic cones using the technique known as "tower of variables" [6]:

$$w^2 \le v_1^{(\rho-1)} v_2^{(\rho-1)} \tag{2.16a}$$

$$\left(v_i^{(l)}\right)^2 \le v_{2i-1}^{(l-1)}\, v_{2i}^{(l-1)}, \quad i = 1, \ldots, 2^{\rho-l}, \quad l = 2, \ldots, \rho - 1, \tag{2.16b}$$

$$\left(v_i^{(1)}\right)^2 \le u^2, \qquad i = 1, \ldots, \lfloor s/2 \rfloor, \tag{2.16c}$$

$$\left(v_i^{(1)}\right)^2 \le ut, \qquad i = \lfloor s/2 \rfloor + 1, \ldots, \lceil s/2 \rceil, \tag{2.16d}$$

$$\left(v_i^{(1)}\right)^2 \le t^2, \qquad i = \lceil s/2 \rceil + 1, \ldots, \lfloor r/2 \rfloor, \tag{2.16e}$$

$$\left(v_i^{(1)}\right)^2 \le tw, \qquad i = \lfloor r/2 \rfloor + 1, \ldots, \lceil r/2 \rceil, \tag{2.16f}$$

$$\left(v_i^{(1)}\right)^2 \le w^2, \qquad i = \lceil r/2 \rceil + 1, \ldots, \lfloor R/2 \rfloor, \tag{2.16g}$$

$$w, v_i^{(\ell)}, u, t \ge 0.$$

The set of inequalities (2.16) can be visualized as a binary tree whose nodes represent the variables in (2.16). Each constraint in (2.16) can then be viewed as a subgraph with two arcs that connect the "parent" node (the variable at the left-hand side of the constraint) to the two "child" nodes (the variables at the right-hand side of the same constraint). Given this binary structure, the set of second-order cones in (2.16) can be regarded as partitioned into $l = 1, \ldots, \rho$ levels, where the variable w in constraint (2.16a) constitutes the root node of the tree, and belongs to ρ-level, while variables u, t, w in (2.16d)–(2.16g) represent the leaf nodes, or 0-level nodes of the tree.

In [28] it has been shown that among the $2^\rho - 1$ inequalities (2.16) there are only

$O(\rho) = O(\log_2 r)$ non-degenerate second-order cones, while the rest reduce to linear inequalities that can be omitted. Further, the following bounds on the number of non-degenerate quadratic cones in (2.16) follow directly from the arguments in [28]:

Proposition 3 (Krokhmal,Soberanis [28]). *When p is a positive rational number, $p = r/s$, such that $r > s$ and the greatest common divisor of r and s is 1, a p-order cone in the positive orthant of $\mathbb{R}^{n+1}$ can equivalently be represented by C_p three-dimensional quadratic cones, where C_p satisfies*

$$n\rho \leq C_p \leq n(2\rho - 1), \quad \rho = \lceil \log_2 r \rceil. \tag{2.17}$$

The proof of Proposition 3 exploits the fact that a non-degenerate quadratic cone constraint in (2.16) corresponds to a subgraph where child nodes have different variables assigned to them, and each level of the tree must necessarily contain at least one such (non-degenerate) constraint, see [28].

It it easy to see that the order in which the variables u, t, and w are assigned to the leaf nodes in the binary tree (2.16) can significantly affect the number of non-degenerate quadratic cones needed to represent a rational-order p-cone in $\mathbb{R}^{n+1}$. As an illustration, consider the case $p = 3$; in accordance to the above we have $\rho = 2$, $R = 4$, and direct application of (2.16) yields a binary tree where the variables u_j, t, and w_j are assigned to the leaf nodes in the order (u_j, t, t, w_j). The resulting representation of $p = 3$ cone (2.13) involves $3n$ three-dimensional rotated quadratic cones:

$$t \geq u_1 + \ldots + u_n; \quad w_j^2 \leq v_{j1}^{(1)} v_{j2}^{(1)}, \quad \left(v_{j1}^{(1)}\right)^2 \leq u_j t, \quad \left(v_{j2}^{(1)}\right)^2 \leq t w_j, \quad j = 1, \ldots, n, \tag{2.18}$$

On the other hand, it is easy to see that an assignment of variables to the leaf nodes in the order (u_j, w_j, t, t) allows for reducing the number of 3D quadratic cones necessary to represent a $p = 3$ cone in $\mathbb{R}^{n+1}_+$ to $2n$:

$$t \geq u_1 + \ldots + u_n; \quad w_j^2 \leq t v_{2j}^{(1)}, \quad \left(v_{2j}^{(1)}\right)^2 \leq u_j w_j, \quad j = 1, \ldots, n. \tag{2.19}$$

Observe that the number of second-order cones in representations (2.18) and (2.19) correspond to the upper and lower bounds in (2.17), respectively.

When the described technique of transforming a p-cone programming problem into SOCP problem is applied in practice, a reduction in the number of second-order cone inequalities in (2.16) leads to a reduction in the number of second-order cone constraints by the order of dimensionality n of the original p-cone (2.13). Hence, it is of interest to devise an "economical" representation of rational-order cones via second-order cones; the next section addresses this issue.

2.3.2 An "economical" representation of rational-order p-cone via second order cones

Clearly, a reduction in the number of second-order cone inequalities in (2.16) leads to a reduction in the number of second-order cone constraints in the optimization problem by the order of dimensionality n of the original p-cone. Below we demonstrate that the lower bound on C_p in (2.17) is achievable for any rational $p \geq 1$, and present an algorithm for constructing the corresponding SOCP representation of a rational-order p-cone. To this end, consider the following convex pointed cone in $\mathbb{R}^4_+$:

$$\mathcal{P} = \left\{ \mathbf{y} \in \mathbb{R}^4_+ \;\middle|\; y_0^{k_0} - y_1^{k_1} y_2^{k_2} y_3^{k_3} \leq 0 \right\}, \tag{2.20}$$

that satisfies the next four properties:

(P1) $k_0, k_1, k_2, k_3 \in \mathbb{Z}_+$;

(P2) $k_0 = k_1 + k_2 + k_3$;

(P3) $k_1 + k_2 + k_3 = 2^q$ for some integer $q \geq 1$;

(P4) exactly two numbers among k_1, k_2, and k_3 are odd.

Proposition 4. *Cone $\mathcal{P}$ (2.20) that satisfies (P1)–(P4) can be represented as an intersection of at most q three-dimensional cones of the form $\left\{ \mathbf{x} \in \mathbb{R}^3_+ \;\middle|\; x_3^2 \leq x_1 x_2 \right\}$.*

Proof. The process of building such a representation of $\mathcal{P}$ is based on successive lifting of $\mathcal{P}$ into spaces of dimensions greater than previous by 1, in such a way that the degree of the polynomial in (2.20) is reduced by half each time.

First, let us assume that $k_1, k_3, k_3 > 0$ are all different and $q \geq 2$. Without loss of generality, let k_1, k_2 be odd and $k_2 > k_1$, and consider the following set in $\mathbb{R}^5_+$:

$$\begin{gathered} \mathcal{P}^* = \left\{ \mathbf{y} \in \mathbb{R}^5_+ \;\middle|\; y_0^{\nu_0} - y_4^{\nu_4} y_2^{\nu_2} y_3^{\nu_3} \leq 0, \;\; y_4^2 \leq y_1 y_2 \right\}, \\ \text{where} \quad \nu_0 = k_0/2, \quad \nu_2 = (k_2 - k_1)/2, \quad \nu_4 = k_1, \quad \nu_3 = k_3/2. \end{gathered} \tag{2.21}$$

It is easy to see that any $(y_0, y_1, y_2, y_3) \in \mathcal{P}$ can be extended to $(y_0, y_1, y_2, y_3, y_4) \in \mathcal{P}^*$, and any $(y_0, y_1, y_2, y_3, y_4) \in \mathcal{P}^*$ is such that $(y_0, y_1, y_2, y_3) \in \mathcal{P}$.

Now, let us check that the first cone of P^* satisfies (P1)–(P4). As k_1 and k_2 are odd and positive integers by assumption, due to (P4) k_3 is even, whence ν_3 is a positive integer. The above assumption also implies that $k_2 - k_1$ is even, means that ν_2 is a positive integer. Similarly, ν_0 is integer and $\nu_0 = 2^{q-1}$. Also, observe that $\nu_1 + \nu_2 + \nu_3 = (k_1 + k_2 + k_3)/2 = k_0/2 = \nu_0$. So, the first cone in (2.21) satisfies properties (P1)–(P3). Next, observe that $\nu_4 = k_1$ is odd, thus out of two integers ν_2, ν_3 exactly one should be odd for $\nu_2 + \nu_3 + \nu_4 = 2^{q-1}$ to hold. Thus, condition (P4) holds as well.

Note that if in our assumption $k_1 = k_2$, then $\nu_2 = 0$ in (2.21), but all conditions still hold. Consider the case when $q \geq 2$ and one of k_1, k_2, k_3 is zero, assume it is k_3. Then k_1, k_2 should be odd by (P4). Performing the same transformation, we obtain

$$\begin{gathered} \mathcal{P}^{**} = \left\{ \mathbf{y} \in \mathbb{R}^5_+ \;\middle|\; y_0^{\nu_0} - y_4^{\nu_4} y_2^{\nu_2} \leq 0, \;\; y_4^2 \leq y_1 y_2 \right\}, \\ \text{where} \quad \nu_0 = k_0/2, \quad \nu_2 = (k_2 - k_1)/2, \quad \nu_4 = k_1. \end{gathered} \tag{2.22}$$

The first cone of $\mathcal{P}^{**}$ still has properties (P1)–(P4), and $(y_0, \ldots, y_3) \in \mathcal{P}$ can be extended to $(y_0, \ldots, y_4) \in \mathcal{P}^{**}$, and any $(y_0, \ldots, y_4) \in \mathcal{P}^{**}$ is such that $(y_0, \ldots, y_3) \in \mathcal{P}$.

If $q = 1$, then one of k_1, k_2, k_3 is zero, and two others are necessarily equal to 1. In this case $\mathcal{P}$ is already a rotated quadratic cone. Thus, the lifting transformation described above can be carried out no more than $q - 1$ times, and the conic set $\mathcal{P}$ (2.20) can be represented by at most q second order cones using at most $q - 1$ new variables.

With the help of Proposition 4 we can now establish the following result on SOCP representation of rational-order p-cones:

Theorem 3. *Let $p > 1$ be a positive rational number, $p = r/s$, where the greatest common divisor of r and s is 1. Then a p-order cone in the positive orthant of $\mathbb{R}^{n+1}$ can equivalently be represented by $n\lceil \log_2 r \rceil$ three-dimensional rotated quadratic cones.*

Proof. In accordance to (2.13)–(2.15), the problem of representing a (r/s)-cone in $\mathbb{R}^{n+1}_+$ via second-order cones can be reduced to finding a second-order cone representation of n sets of the form

$$\mathcal{Q} = \left\{ \mathbf{y} \in \mathbb{R}^3_+ \;\middle|\; y_3^R - y_1^s y_2^{r-s} y_3^{R-r} \leq 0 \right\}, \tag{2.23}$$

where $R = 2^{\rho}$, $\rho = \lceil \log_2 r \rceil$. Observe that cone $\mathcal{Q}$ is equivalent to intersection of cone $\mathcal{P}$ (2.20), where $k_1 = s$, $k_2 = r - s$, $k_3 = R - r$, with a hyperplane $y_0 = y_3$. Indeed, properties (P1)–(P3) are obvious, and (P4) holds since if r and s do not have common divisor greater than 1, neither do $r - s$ and s, whereby $r - s$ and s cannot be both even.

Note that an iteration of the lifting procedure described in Proposition 4 corresponds to a specific order in which the variables at some level of the binary tree are arranged. For example, the first iteration of lifting corresponds to arranging the 0-level variables $\{w, t, u\} = \{y_1, y_2, y_3\}$ in pairs corresponding to second-order cone constraints, such that y_1 and y_2 make k_1 pairs, or $y_4^2 \leq y_1 y_2$ non-degenerate cones; the remaining $k_2 - k_1$ variables y_2 form $(k_2 - k_1)/2$ pairs, or degenerate cones ${y'_4}^2 \leq y_2^2$, and k_3 variables y_3 form $k_3/2$ pairs, or degenerate cones ${y''_4}^2 \leq y_3^3$, assuming that $k_1 < k_2$ are odd. Obviously, the degenerate cones can simply be disregarded.

Hence, by Proposition 4, $\mathcal{Q}$ admits representation by at most $\rho = \lceil \log_2 r \rceil$ second order cones; combining this with Proposition 3, one obtains that each of n sets of the form $\mathcal{Q}$ admits representation using exactly $\rho = \lceil \log_2 r \rceil$ second order cones.

The lifting procedure outlined in the proof of Proposition 4 can be used to construct SOCP representations of rational order p-cones. The procedure is formalized in Algorithm 1.

It is well known that second order cone constrains admit an equivalent semidefinite representation in the form of linear matrix inequalities (LMIs). In general, p-order cones are not LMI-representable in the space of original variables (see an example for $p = 4$ cone in [20, 21]), but admit lifted LMI representations.

Algorithm 2.1 Reduction of cone $\mathcal{P}$ (2.20) to a set of 3D second-order cones

Input: Cone $\mathcal{P} = \{\mathbf{y} \in \mathbb{R}^4_+ \mid y_0^{k_0} \leq y_1^{k_1} y_2^{k_2} y_3^{k_3}\}$

Output: Set $\mathcal{S}$ of three-dimensional quadratic cone constraints of the form $y_i^2 \leq y_{i_m} y_{i_n}$, defined on the set of variables $\mathcal{y}$.

$\mathcal{S} := \emptyset$;

$\mathcal{y} := \{y_0, y_1, y_2, y_3\}$;

$i := 4$; // counter of a new variable to be added;

$j := 1$; // step counter;

$k_\nu^{(j)} := k_\nu$, $\nu = 0, \ldots, 3$;

$\{l, m, n\} \leftarrow \{1, 2, 3\}$ such that $k_l^{(j)}$ is even and $k_m^{(j)} \geq k_n^{(j)}$ are odd;

while $k_m^{(j)} + k_n^{(j)} + k_l^{(j)} > 2$ **do**

add new variable $\mathcal{y} := \mathcal{y} \cup \{y_i\}$;

add cone: $\mathcal{S} := \mathcal{S} \cup \{y_i^2 \leq y_m y_n\}$;

$k_0^{(j+1)} := k_0^{(j)}/2$; $k_l^{(j+1)} := k_l^{(j)}/2$; $k_m^{(j+1)} := (k_m^{(j)} - k_n^{(j)})/2$; $k_i^{(j+1)} := k_n^{(j)}$;

update $\{l, m, n\} \leftarrow \{l, m, i\}$ such that $k_l^{(j+1)}$ is even and $k_m^{(j+1)} \geq k_n^{(j+1)}$ are odd;

$i := i + 1, j := j + 1$;

end while

add cone: $\mathcal{S} := \mathcal{S} \cup \{y_0^2 \leq y_m y_n\}$;

Corollary 1. *Conic set* $\mathcal{Q}$ (2.23) *admits a lifted representation in the form of LMI*

$$\mathcal{Q}^* = \left\{ \mathbf{y} \in \mathbb{R}^{\rho+2}_+ \;\middle|\; \sum_{i=1}^{\rho+2} \mathbf{A}_i y_i \succeq \mathbf{0} \right\}, \tag{2.24}$$

where $\mathbf{A}_i \in \mathbb{R}^{2\rho \times 2\rho}$ *are symmetric matrices, in the sense that the projection of* $\mathcal{Q}^*$ *onto the space of variables* (y_1, y_2, y_3) *coincides with* $\mathcal{Q}$.

2.4 Computational study

In this section we report computational results on using the p-norm discrimination model (2.6) for linear separation of sets. Recall that the p-norm linear separation model (2.6) can be presented in the form of p-order cone programming problem (pOCP) (2.7) with two p-order cone constraints. Below we illustrate the SOCP reformulation approach to solving (2.7) in the case when p is rational, and compare it with the polyhedral approximation technique of [28].

2.4.1 SOCP reformulation

Applying the approach delineated in Section 3, in the case of a rational $p = r/s$ the pOCP problem (2.7) can be reformulated as a SOCP problem

$$\min \quad \delta_1 k^{-1/p}\, \xi + \delta_2 m^{-1/p}\, \eta \tag{2.25a}$$

$$\text{s.t.} \quad \text{(2.7d), (2.7e)}, \tag{2.25b}$$

$$\xi \geq \mathbf{e}^\top \mathbf{u}_y, \tag{2.25c}$$

$$\eta \geq \mathbf{e}^\top \mathbf{u}_z, \tag{2.25d}$$

$$(\mathbf{y}, \mathbf{u}_y, \mathbf{v}_y, \xi) \in \mathcal{S}_k^{(r/s)} \subset \mathbb{R}_+^{k\lceil \log_2 r \rceil + k + 1} \tag{2.25e}$$

$$(\mathbf{z}, \mathbf{u}_z, \mathbf{v}_z, \eta) \in \mathcal{S}_m^{(r,/s)} \subset \mathbb{R}_+^{m\lceil \log_2 r \rceil + m + 1} \tag{2.25f}$$

$$\mathbf{y}, \mathbf{z}, \mathbf{u}_y, \mathbf{u}_z, \eta, \xi \geq 0, \tag{2.25g}$$

where $\mathcal{S}_n^{(r/s)}$ denotes the set of three-dimensional second-order cones generated by Algorithm 2.1 to represent a $(n+1)$-dimensional (r/s)-order cone. The SOCP reformulation (2.25) allows for employing efficient self-dual optimization methods for solving the p-norm separation model (2.7), but this comes at the expense of a large number of second-order cones in (2.25). In order to benchmark the efficiency of such an approach, we compare it with the solution method that is based on solving polyhedral approximations of pOCP problems using a cutting plane technique [28].

2.4.2 A polyhedral approximation procedure

The polyhedral approximation of pOCP problems obtained in [28] relies on representing p-order cones in $\mathbb{R}^{n+1}$ as intersection of 3-dimensional p-cones using a proce-

dure similar to that described in Section 2.3.1. Assuming, for expositional simplicity, that $n = 2^d$ for some integer d, the p-cone (2.13) can be represented as

$$\begin{aligned} t &\geq \left\| \left(w_1^{(d-1)}, w_2^{(d-1)} \right) \right\|_p, \\ w_j^{(l)} &\geq \left\| \left(w_{2j-1}^{(l-1)}, w_{2j}^{(l-1)} \right) \right\|_p, \quad j = 1, \ldots, 2^{d-l}, \quad l = 2, \ldots, d-1, \\ w_j^{(1)} &\geq \left\| \left(w_{2j-1}, w_{2j} \right) \right\|_p, \quad j = 1, \ldots, 2^{d-1} \end{aligned} \tag{2.26}$$

Then, each 3D p-cone $z \geq \|(x, y)\|_p$ in (2.26) is replaced by by its outer polyhedral approximation via $m + 1$ tangent planes

$$z \geq \alpha_i^{(p)}(m)\, x + \beta_i^{(p)}(m)\, y, \quad i = 0, \ldots, m, \tag{2.27}$$

where the coefficients $\alpha_i^{(p)}(m)$, $\beta_i^{(p)}(m)$ depend on the parameter of construction m that controls approximation accuracy:

$$\alpha_i^{(p)}(m) = (\cos^p \theta_i + \sin^p \theta_i)^{\frac{1-p}{p}} \cos^{p-1} \theta_i, \quad \beta_i^{(p)}(m) = (\cos^p \theta_i + \sin^p \theta_i)^{\frac{1-p}{p}} \sin^{p-1} \theta_i, \quad \theta_i = \frac{\pi i}{2m}.$$

When applied to the p-cone programming model (2.7), this polyhedral approximation technique allows for replacing two p-cone constraints (2.7b)–(2.7c) with $m_y(2^d - 1) + m_z(2^h - 1)$ linear constraints, where the parameters m_y, m_z determine the number of facets in polyhedral approximations of 3D p-cones (2.26) corresponding to p-cones (2.7b) and (2.7c), respectively, necessary to achieve the prescribed approximation accuracy ε. Following [15], the number of approximating linear constraints can be reduced by allowing the number of facets in polyhedral approximations (2.27) of 3D p-cones (2.26) to vary with l, such that the total number of facets used for approximation of the high-dimensional p-cone is minimized while guaranteeing the prescribed approximation accuracy ε:

$$\min \left\{ \sum_{l=1}^{t} q_l m_l \;\middle|\; 1 + \varepsilon \geq \prod_{l=1}^{t} (1 + \epsilon_l(m_l)), \;\; m_l \in \mathbb{Z}_+ \right\}, \tag{2.28}$$

where, for a given l, m_l is the number of facets in the polyhedral approximation (2.27) of a 3D p-cone, and $\epsilon_l = \epsilon_l(m_l)$ is the main term of the corresponding approximation accuracy [28]:

$$\epsilon_l(m_l) = \begin{cases} \frac{1}{p}(1-\frac{1}{p})^p\left(\frac{\pi}{2m_l}\right)^p, & p \in (1,2) \\ \\ \frac{1}{8}(p-1)\left(\frac{\pi}{2m_l}\right)^2, & p \in [2,\infty), \end{cases}$$

Problem (2.28) can be solved by rewriting its constraint in a logarithmic form: $\sum_{l=1}^{t} \ln(1+\epsilon_l(m_l)) \leq \ln(1+\varepsilon)$, which in turn can be replaced, due to the inequality $\ln(1+x) \leq x$, with $\sum_{l=1}^{t} \epsilon_l(m_l) \leq \ln(1+\varepsilon)$. Then, relaxation of the resulting nonlinear integer programming problem was solved using the method of Lagrange multipliers. For simplicity, we let $m_l = \lceil m_l^* \rceil$, where m_l^* is the solution of relaxed problem. This procedure resulted in, on average, a 50% reduction in the number of approximating facets. The resulting linear programming problem was then solved using the cutting plane procedure described in [28].

2.4.3 SVM analogy

Support Vector Machine is widely used in classification problems. In the general case of linear non-separable sets SVM method can be written in the form of quadratic optimization problem as follows:

$$\min \quad \frac{1}{2}\|\mathbf{w}\|_2 + \sum_{i=1}^{m}\sum_{j=1}^{k} C_1\varepsilon_i^a + C_2\varepsilon_j^b \tag{2.29a}$$

$$\text{s. t.} \quad \mathbf{w}^T\mathbf{a}_i - \gamma \geq 1 - \varepsilon_i^a, \quad i = 1\ldots m \tag{2.29b}$$

$$-\mathbf{w}^T\mathbf{b}_j + \gamma \geq 1 - \varepsilon_j^b, \quad j = 1\ldots k \tag{2.29c}$$

$$\varepsilon_i^a \geq 0, \quad i = 1\ldots m \tag{2.29d}$$

$$\varepsilon_j^b \geq 0, \quad j = 1\ldots k \tag{2.29e}$$

where C_1,C_2 are positive constants.

Note that $(\mathbf{w}^T\mathbf{a}_i - \gamma)$ and $(-\mathbf{w}^T\mathbf{b}_j + \gamma)$ give the misclassification errors, so (2.29) can be rewritten in terms of vectors $\mathbf{y}$ and $\mathbf{z}$ (C_1, C_2 are positive constants):

$$\min\left\{\frac{1}{2}\|\mathbf{w}\|_2 + C_1\,\xi + C_2\,\eta \;\middle|\; \xi \geq \|\mathbf{y}\|_1, \;\; \eta \geq \|\mathbf{z}\|_1, \;\; (2.7d), \;\; (2.7e), \;\; (2.7f)\right\}, \tag{2.30}$$

Given $C_1 = \delta_1/k$, $C_2 = \delta_2/m$ and recalling formulation of the p-norm discrimination problem (2.7), one can conclude that

$$C_1\,\xi^*_{(p)} + C_2\,\eta^*_{(p)} \leq \frac{1}{2}\|\mathbf{w}^*_{(SVM)}\|_2 + C_1\,\xi^*_{(SVM)} + C_2\,\eta^*_{(SVM)} \leq \frac{1}{2}\|\mathbf{w}^*_{(p)}\|_2 + C_1\,\xi^*_{(p)} + C_2\,\eta^*_{(p)}$$

where $(w^*_{(SVM)}, \xi^*_{(SVM)}, \eta^*_{(SVM)})$ stands for the optimal solution of the SVM problem (2.30), and $(w^*_{(p)}, \xi^*_{(p)}, \eta^*_{(p)})$ refers to the optimal solution of p-norm discrimination problem (2.7). Indeed, as $\|w_{(SVM)}\|_2$ is non-negative and $(w^*_{(SVM)}, \xi^*_{(SVM)}, \eta^*_{(SVM)})$,$(w^*_{(p)}, \xi^*_{(p)}, \eta^*_{(p)})$ are optimal solutions of corresponding problems, the following inequality holds:

$$\frac{1}{2}\|\mathbf{w}^*_{(SVM)}\|_2 + C_1\,\xi^*_{(SVM)} + C_2\,\eta^*_{(SVM)} \geq C_1\,\xi^*_{(SVM)} + C_2\,\eta^*_{(SVM)} \geq$$

$$\geq C_1\,\xi^*_{(1)} + C_2\,\eta^*_{(1)} \geq C_1\,\xi^*_{(p)} + C_2\,\eta^*_{(p)}.$$

Hence, the optimal objective value of SVM problem (2.30) gives upper bound for the optimal objective value of p-norm discrimination problem (2.7).

2.4.4 Computational results

In our computational experiments we used three datasets from UCI Machine Learning Repository. The first dataset is Wisconsin Breast Cancer Dataset with a total of 683 instances and 9 attributes. It contains 444 instances with benign diagnosis (type A) and 239 instances with malignant diagnosis (type B). The second dataset, Cleveland Heart Disease Dataset, contains 281 instances with 13 attributes, of them 125 instances correspond to positive diagnosis and 156 instances correspond to negative diagnosis. Finally, the Pima Indians Diabetes Dataset reports 768 instances with 8 attributes, including 266 instances of positive diagnosis and 502 instances of negative diagnosis. Both the Wisconsin Breast Cancer and Cleveland Heart Disease datasets (in their then-up-to-date versions) were used in [8], and can be regarded as relatively well suited for linear discrimination methods; in contrast, the Pima Indians dataset appears to be less suitable for linear separation.

For each dataset, training and testing was performed by randomly selecting 100 training sets with equal number of points of both types, and testing the obtained separator on the data not included in the training set. For computational purposes, the data in training datasets was normalized and scaled by a factor of 10^4; the same transformation was then applied to testing data. After the training and testing procedures were performed, the average misclassification error on testing set was computed. For each value of the parameter p in the range of 1 to 4 (with a step of 0.1), the corresponding p-norm separating problem

(2.7) with $\delta_{1,2} = 1$ was solved using interior-point SOCP solver via SOCP reduction described in Section 3.1. In addition, a polyhedral approximation of (2.7) was solved using the cutting plane procedure of [28]. The corresponding SOCP and LP optimization models were implemented in C++ and solved using IBM CPLEX 12.2 solver.

Table 2.1 reports the smallest average out-of-sample misclassification error for each dataset, together with the corresponding value of p at which this error was obtained, and compares it with the misclassification error for the case of $p = 1$ (which corresponds to the method proposed due to 8). Figures 2.1, 2.2, and 2.3 illustrate the behavior of the misclassification error with respect to the value of parameter p in (2.6) for the described datasets. As it follows from Table 2.1 and Figures 2.1–2.3, the p-norm separation model (2.6)–(2.7) with $p > 1$ allows for an improved classification accuracy as compared to the cases of $p = 1$ proposed in [8] and SVM method.

In addition to classification capabilities of the p-norm linear separation model (2.6)–(2.7), its computational properties were investigated. In particular, for all the datasets described above we compared the running times of the cutting plane procedure for polyhedral approximation of problem (2.7), denoted as LP/CP, and the "economical" SOCP reformulation of (2.7) solved by CPLEX's barrier solver (denoted as SOCP). All computations were performed on a dual-core 3GHz CPU computer with 2GB of RAM. In addition to the running times of the LP/CP and SOCP algorithms, Figures 2.4, 2.5, 2.6 display the values of the parameter $\rho = \log_2 r$, where $p = r/s$, that is proportional to the number of second-order cones in the SOCP reformulation of rational-order p-cone programming problem (2.7). From Figures 2.4–2.6 it follows that the solution times for SOCP reformulation

of a rational-order p-cone programming model (2.7) is highly correlated to the number of second-order cones in the reformulated problem. On the other hand, the solution times of a polyhedral approximation of (2.7) solved with a cutting plane method (LP/CP) exhibit relatively little dependence on the value of the parameter p.

Table 2.1: Classification results for different datasets: the lowest average misclassification error, the corresponding value of p, and misclassification error for the case of $p = 1$, which corresponds to the method proposed in Bennett, Mangasarian (1992)

Dataset	Error	Best p	$p = 1$
Wisconsin Breast Cancer Dataset	3.95%	1.8	4.11%
Cleveland Heart Disease Dataset	18.7%	3.8	19.5%
Pima Indians Diabetes Dataset	31.82%	3.4	35.29%

Figure 2.1: Misclassification error as a function of p for Wisconsin Breast Cancer dataset

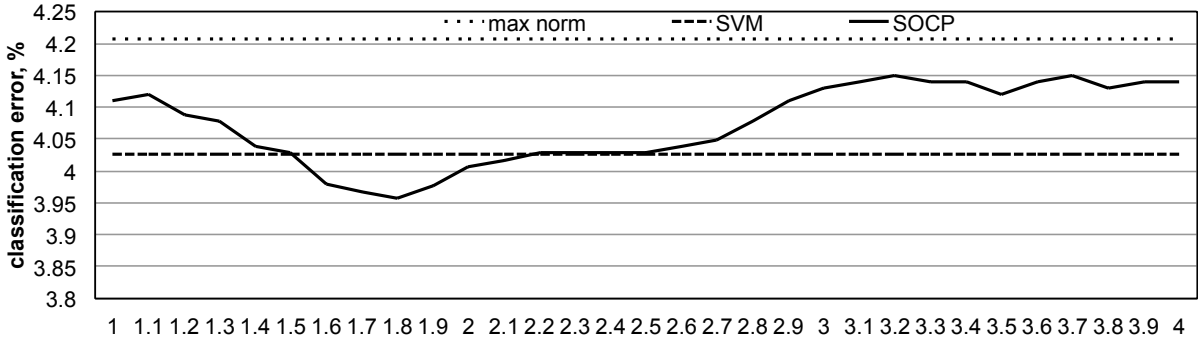


Figure 2.2: Misclassification error as a function of p for Cleveland Heart Disease dataset

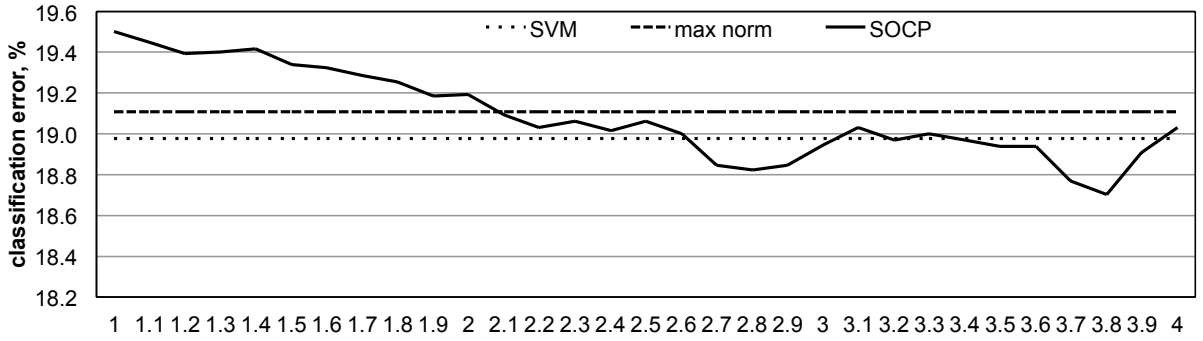


Figure 2.3: Misclassification error as a function of p for Pima Indians Diabetes dataset

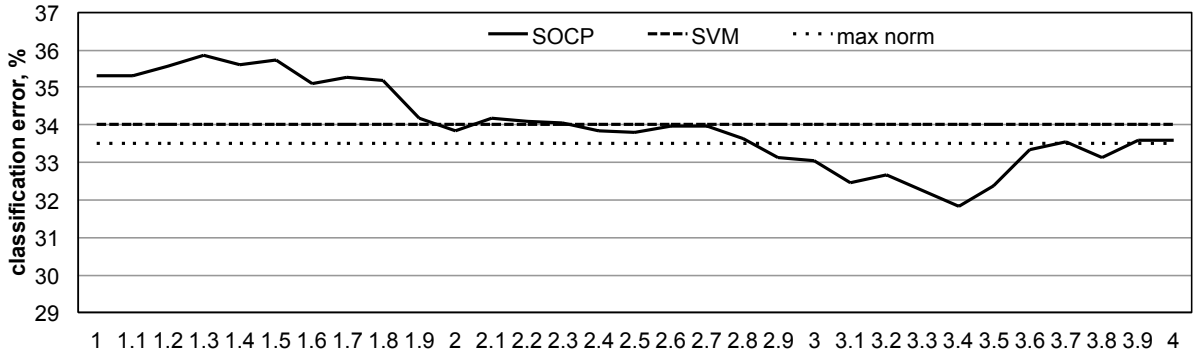


Table 2.2: Comparison of Running Time for Cleveland Heart Disease Dataset: LP/CP stands for cutting plane approximation method, SOCP denotes running time for CPLEX solver on the initial problem (2.4) using Second Order Conic Representation

p	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
SOCP	0.000	2.538	1.890	2.647	1.944	1.214	1.733	3.089	2.476	3.105
LP/CP	0.000	0.113	0.154	0.186	0.215	0.241	0.265	0.295	0.316	0.323
p	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9
SOCP	0.787	3.146	2.583	3.334	2.583	1.857	2.283	3.702	2.385	3.015
CG	0.341	0.368	0.421	0.412	0.421	0.459	0.481	0.491	0.511	0.508
p	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
SOCP	1.177	3.238	2.245	4.457	2.825	1.813	2.958	3.942	2.968	4.314
LP/CP	0.530	0.541	0.540	0.553	0.562	0.592	0.610	0.601	0.594	0.598
p	4.0									
SOCP	1.130									
LP/CP	0.590									

Figure 2.4: Running time comparison of LP/CP and SOCP solution methods of the p-norm separation problem for the Wisconsin Breast Cancer Dataset. The value of parameter ρ determines the number of second-order cones in the SOCP reformulation of problem (2.7)

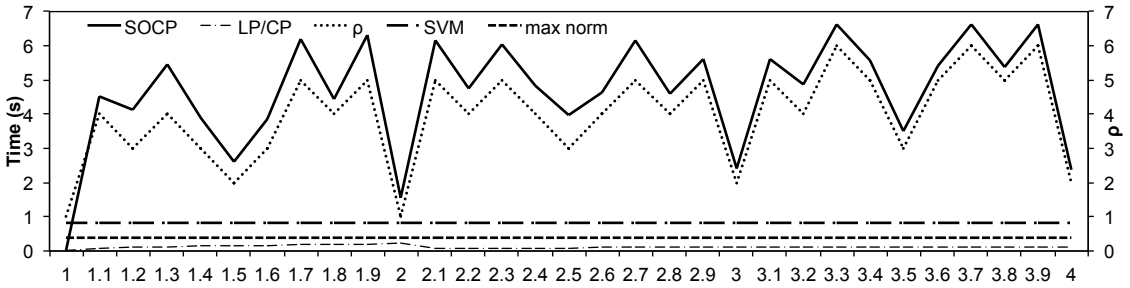


Figure 2.5: Running time comparison of LP/CP and SOCP solution methods of the p-norm separation problem for the Cleveland Heart Disease Dataset. The value of parameter ρ determines the number of second-order cones in the SOCP reformulation of problem (2.7)

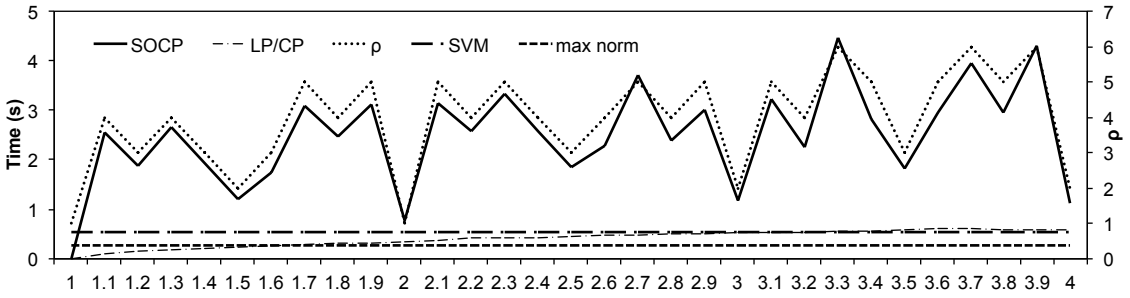


Table 2.3: Comparison of Running Time for Pima Indians Diabetes Dataset: LP/CP stands for cutting plane approximation method, SOCP denotes running time for CPLEX solver on the initial problem (2.4) using Second Order Conic Representation

p	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
SOCP	0.127	4.217	3.847	4.895	3.598	2.4295	3.498	5.664	4.434	5.379
LP/CP	0.104	0.104	0.148	0.173	0.194	0.205	0.214	0.225	0.233	0.244
p	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9
SOCP	1.395	5.904	4.506	5.314	4.570	3.303	4.105	5.251	4.159	5.264
LP/CP	0.253	0.258	0.256	0.255	0.253	0.250	0.256	0.259	0.260	0.256
p	3	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
SOCP	2.111	5.164	4.299	6.006	4.944	3.067	4.686	6.214	4.737	5.915
CG	0.260	0.267	0.251	0.261	0.256	0.261	0.261	0.255	0.258	0.258
p	4.0									
SOCP	2.159									
LP/CP	0.252									

Figure 2.6: Running time comparison of LP/CP and SOCP solution methods of the p-norm separation problem for the Pima Indians Diabetes Dataset. The value of parameter ρ determines the number of second-order cones in the SOCP reformulation of problem (2.7)

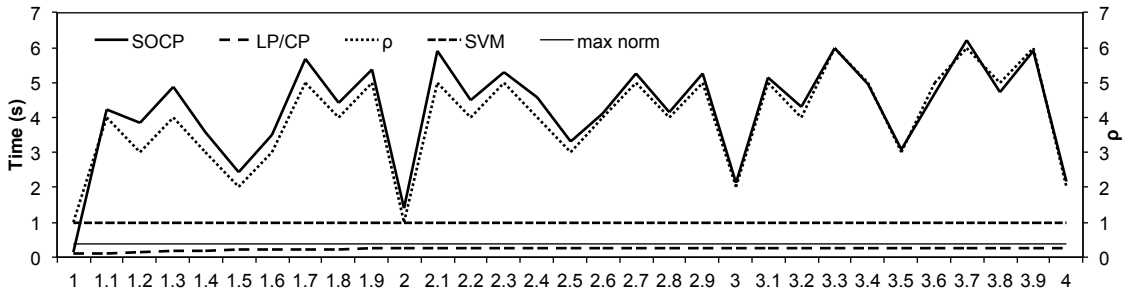


Table 2.4: Comparison of Running Time for Wisconsin Breast Cancer Dataset: LP/CP stands for cutting plane approximation method, SOCP denotes running time for CPLEX solver on the initial problem (2.4) using Second Order Conic Representation

p	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
SOCP	0.016	4.534	4.131	5.453	3.881	2.6218	3.849	6.201	4.439	6.312
LP/CP	0.016	0.084	0.103	0.153	0.136	0.153	0.167	0.181	0.196	0.205
p	2	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9
SOCP	1.575	6.150	4.770	6.051	4.823	3.980	4.631	6.158	4.602	5.622
LP/CP	0.219	0.184	0.187	0.189	0.091	0.095	0.097	0.099	0.102	0.105
p	3	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
SOCP	2.419	5.598	4.855	6.623	5.580	3.4904	5.408	6.612	5.394	6.634
LP/CP	0.107	0.108	0.109	0.114	0.114	0.119	0.120	0.118	0.122	0.123
p	4									
SOCP	2.386									
LP/CP	0.124									

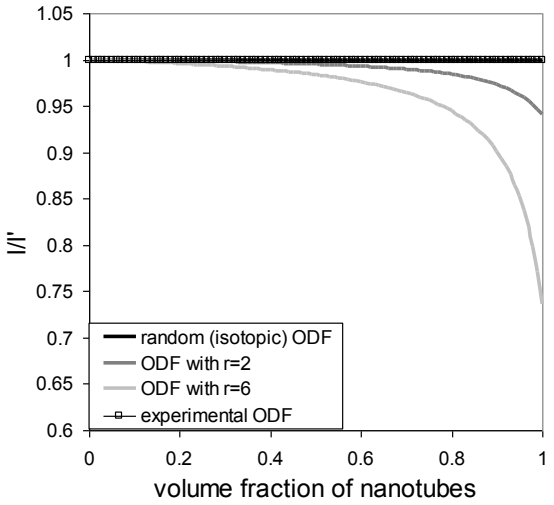
CHAPTER 3
SEMIDEFINITE PROGRAMMING MODELS FOR DETERMINING BOUNDS ON THE OVERALL PROPERTIES OF COMPOSITE MATERIALS WITH RANDOMLY ORIENTED INCLUSIONS

3.1 Introduction

The main motivation for this work comes from study of the effects of the orientational distribution of carbon nanotubes (CNTs) in buckypapers on the overall elastic properties of CNT buckypaper polymer matrix composites. Buckypapers are thin sheets of porous carbon nanotubes networks that are prepared by a multi-step process of dispersion and filtration of nanotube suspension. Bulk buckypaper polymeric composites are obtained by impregnation of nanotube buckypapers into a polymer matrix. Unless a special care is taken, the nanotubes are distributed randomly in buckypaper sheets. To achieve certain alignment, buckypaper sheets are produced by filtrating well-dispersed nanotube suspension through a filter placed in a high strength magnetic field [29]. A strong magnetic field (5-15 T) substantially improves alignment of CNTs and thereby increases the buckypaper's elastic modulus and strength in the direction of alignment. The alignment is described by the orientation distribution function. Orientation distribution function (ODF) is used to describe CNT orientation distributions in buckypaper. Figure 1 shows ODF derived from the analysis of the SEM images of CNT buckypapers. ODFs are routinely included in the micromechanical analysis. At the same time, there are no rigorous bounds derived for the composites with orientational distribution (except for the random uniform distribution) of phases. Thus, the validity of the results of the micromechanical analysis cannot be

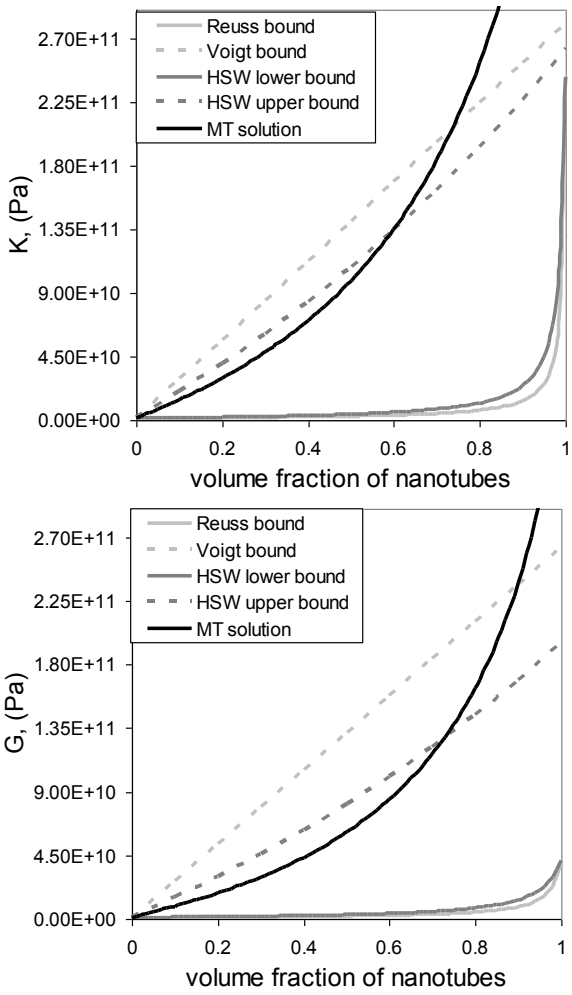
established. Moreover, it has been shown that the Mori-Tanaka scheme applied to the non-aligned CNT buckypaper polymer matrix composites leads to violation of symmetry of the effective elastic moduli tensor (Figure 3.1) and produces the results outside of bounds (Figure 3.2). An extensive discussion of these issues can be found in the work of Zhupanska [46].

Figure 3.1: The ratio l/l' obtained using the MT approach for different degrees of the nanotube alignment ("random (isotropic) ODF" and "experimental ODF" lines do coincide)



To derive the tightest bounds for the composites with non-aligned phases, we proposed to formulate a problem as a nonlinear semidefinite optimization problem, i.e., an optimization problem where the optimization variables are represented by symmetric positive semidefinite matrices. Such a formulation guarantees that any solution of the optimization problem represents a valid tensor of elastic material properties and preserves the symmetry

Figure 3.2: Effective elastic moduli for randomly oriented nanotubes



of the tensor. The problem is then solved by an interior point method to find bounds for the case of random (uniform) distribution of fibers in the matrix.

3.2 Orientation distribution function

Orientation distribution function, ODF, describes orientational distributions of fibers in the matrix and is defined as an orientation probability density. Thus, ODF must

satisfy the following equality:

$$\int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi} F(\phi,\theta,\psi)\sin\theta\; d\theta d\phi d\psi = 1 \tag{3.1}$$

where $F(\phi,\theta,\psi)$ is the ODF and ϕ,θ,ψ are the Euler angles. The random (uniform) distribution of the fibers is described by the following ODF $F(\phi,\theta,\psi) = 1/8\pi^2$. ODFs enter a micromechanical analysis through averaging of the corresponding material properties tensors over all possible directions.

The following orientation distribution functions are considered in this work:

$$F_1(\theta) = \frac{r+1}{8\pi^2}\cos^r(\theta).$$

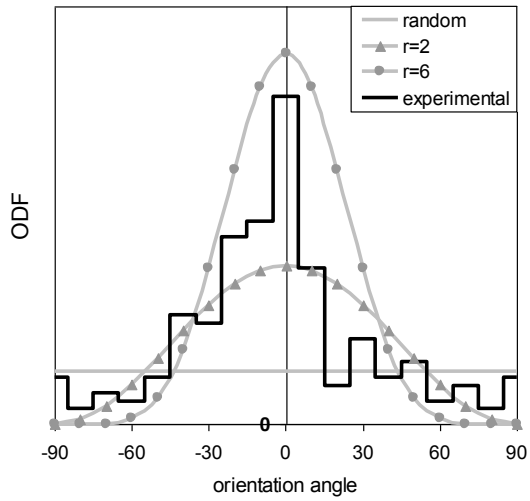
Here the parameter r determines the degree of alignment of inclusions (i.e., nanotubes) with respect to the X_1-axis ($\theta = 0$ along the X_1-axis) and is an even integer number. As r increases the degree of inclusions alignment also increases. If, for instance, $r = 0$, then $F_1(\theta) = 1/8\pi^2$ describes random (uniform) distribution of nanotubes. Figure 3.3 shows theoretical orientation distribution functions for $r = 0$, $r = 2$, $r = 6$, and an experimentally measured ODF.

If the symmetry is preserved, then as a result of computations using (3), or, in other words, .

3.3 Averaging for the overall elastic properties

In the notations of Hill [22] and Walpole [41] the tensor of elastic moduli, L_i, of the i-th phase is $L_i = (2k_i, l_i, l_i', n_i, 2m_i, 2p_i)$ and the tensor of overall elastic moduli is written as $L = (2k, l, l', n, 2m, 2p)$. Both tensors L and L_i have to posses a diagonal symmetry:

Figure 3.3: Orientation distribution functions



$l = l'$, $l_i = l'_i$. In conjunction with the used notations, Hooke's law for the composite with transversely isotropic properties (X_1 is axis of material symmetry) takes the form

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{bmatrix} = \begin{bmatrix} n & l & l & 0 & 0 & 0 \\ l' & m+k & k-m & 0 & 0 & 0 \\ l' & k-m & m+k & 0 & 0 & 0 \\ 0 & 0 & 0 & 2m & 0 & 0 \\ 0 & 0 & 0 & 0 & 2p & 0 \\ 0 & 0 & 0 & 0 & 0 & 2p \end{bmatrix} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ \varepsilon_{23} \\ \varepsilon_{13} \\ \varepsilon_{12} \end{bmatrix}$$

Here k, l, n, m and p are plain bulk modulus, transverse cross modulus and axial modulus under a uniaxial strain, the transverse shear and the axial shear moduli, respectively.

Recall that according to Euler's rotation theorem, any rotation can be described by only three parameters, namely, the Euler angles ϕ, θ, and ψ. Let (x_1, x_2, x_3) to be a local coordinate system and (X_1, X_2, X_3) to be a global coordinate system, then an arbitrary fourth-order tensor B possesses the following transformation rule

$$B^X_{klmn} = \sum_{p,q,s,t} a_{kp} a_{lq} a_{ms} a_{nt} B^x_{pqst}, \tag{3.2}$$

where B^X_{klmn} and B^x_{pqst} are components of the tensor B in the global and local coordinate systems, respectively, and a_{ij} are functions of the Euler angles:

$$\begin{aligned}
a_{11} &= \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi,\\
a_{12} &= \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi,\\
a_{13} &= \sin\psi\sin\theta,\\
a_{21} &= -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi,\\
a_{22} &= -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi,\\
a_{23} &= \cos\psi\sin\theta,\\
a_{31} &= \sin\theta\sin\phi, \quad a_{32} = -\sin\theta\cos\phi, \quad a_{33} = \cos\theta.
\end{aligned} \tag{3.3}$$

Introducing the orientation distribution function, $F(\phi,\theta,\psi)$, we can rewrite the orientationally averaged tensor B in the global coordinate system (X_1, X_2, X_3) as

$$\begin{aligned}
\left\{B^X_{klmn}\right\} &= \int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi} F(\phi,\theta,\psi) B^X_{klmn}\sin\theta\; d\theta d\phi d\psi\\
&= \int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi} F(\phi,\theta,\psi)\sum_{p,q,s,t}\left(a_{kp}a_{lq}a_{ms}a_{nt}B^x_{pqst}\right)\\
&\quad\times\sin\theta\; d\theta d\phi d\psi
\end{aligned} \tag{3.4}$$

Next we provide the explicit formula for transformation (3.2) of transversely isotropic tensors. In the case of an arbitrary fourth-order tensor B the transformation (3.2) is

$$B^X_{klmn} = \sum_{p,q,s,t} a_{kp}a_{lq}a_{ms}a_{nt}B^x_{pqst}$$

where B^X_{klmn} are components of the tensor B in the global coordinate system (X_1, X_2, X_3) and B^x_{pqst} are components of the tensor B in the local coordinate system (x_1, x_2, x_3), and a_{ij} are functions of the Euler angles.

Let us consider a transversely isotropic non-symmetric tensor B

$$B = \begin{bmatrix} d & g & g & 0 & 0 & 0 \\ h & \frac{c+e}{2} & \frac{c-e}{2} & 0 & 0 & 0 \\ h & \frac{c-e}{2} & \frac{c+e}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & e & 0 & 0 \\ 0 & 0 & 0 & 0 & f & 0 \\ 0 & 0 & 0 & 0 & 0 & f \end{bmatrix}$$

In our case, $B = (L_0^* + L_1)^{-1}$ and B is symmetric, $h = g$. Tensor B can be written in a more compact form may be written as $B = (c, g, h, d, e, f)$, or as a vector

$$\mathbf{b} = (c, g, h, d, e, f)^\top$$

Then six components $c^X, g^X, h^X, d^X, e^X, f^X$ of the tensor B in the global coordinates are linked to the six components $c^x, g^x, h^x, d^x, e^x, f^x$ in the local coordinates by

$$\mathbf{b}^X = T\mathbf{b}^x,$$

where vector $\mathbf{b}^X = (c^X, g^X, h^X, d^X, e^X, f^X)^\top$ is the vector $\mathbf{b}$ in the global coordinates, $\mathbf{b}^x = (c^x, g^x, h^x, d^x, e^x, f^x)^\top$ is the vector $\mathbf{b}$ in the local coordinates, and the transformation matrix T is

$$T = \begin{bmatrix} C_c & C_g & C_h & C_d & C_e & C_f \\ G_c & G_g & G_h & G_d & G_e & G_f \\ H_c & H_g & H_h & H_d & H_e & H_f \\ D_c & D_g & D_h & D_d & D_e & D_f \\ E_c & E_g & E_h & E_d & E_e & E_f \\ F_c & F_g & F_h & F_d & F_e & F_f \end{bmatrix}$$

components of which are functions of the Euler angles and expressible as:

$$C_c = \tfrac{1}{2}\left(a_{22}^2 + a_{23}^2\right)\left(a_{22}^2 + a_{23}^2 + a_{32}^2 + a_{33}^2\right),$$

$$C_g = a_{21}^2\left(a_{22}^2 + a_{23}^2 + a_{32}^2 + a_{33}^2\right),$$

$$C_e = 2a_{22}a_{23}a_{32}a_{33} + \tfrac{1}{2}\left(a_{22}^2 + a_{23}^2\right)^2 + \tfrac{1}{2}\left(a_{22}^2 - a_{23}^2\right)\left(a_{32}^2 - a_{33}^2\right),$$

$$C_h = \left(a_{21}^2 + a_{31}^2\right)\left(a_{22}^2 + a_{23}^2\right), \quad C_d = a_{21}^2\left(a_{21}^2 + a_{31}^2\right),$$

$$C_f = 2\left[a_{21}^2\left(a_{22}^2 + a_{23}^2\right) + 2a_{31}a_{32}a_{21}a_{22}\right],$$

$$G_c = \tfrac{1}{2}\left(a_{32}^2 + a_{33}^2\right)\left(a_{12}^2 + a_{13}^2\right), \quad G_d = a_{11}^2 a_{31}^2,$$

$$G_g = a_{11}^2\left(a_{32}^2 + a_{33}^2\right),$$

$$G_e = 2a_{12}a_{13}a_{32}a_{33} + \tfrac{1}{2}\left(a_{12}^2 - a_{13}^2\right)\left(a_{32}^2 - a_{33}^2\right),$$

$$G_h = a_{31}^2\left(a_{12}^2 + a_{13}^2\right), \quad G_f = 2a_{11}a_{31}\left(a_{12}a_{32} + a_{13}a_{33}\right),$$

$$H_c = \tfrac{1}{2}\left(a_{22}^2 + a_{23}^2\right)\left(a_{12}^2 + a_{13}^2\right), \quad H_d = a_{11}^2 a_{21}^2,$$

$$H_g = a_{21}^2\left(a_{12}^2 + a_{13}^2\right),$$

$$H_e = 2a_{12}a_{13}a_{22}a_{23} + \tfrac{1}{2}\left(a_{12}^2 - a_{13}^2\right)\left(a_{22}^2 - a_{23}^2\right),$$

$$H_h = a_{11}^2\left(a_{23}^2 + a_{22}^2\right), \quad H_f = 2a_{11}a_{21}\left(a_{12}a_{22} + a_{13}a_{23}\right),$$

$$D_c = \tfrac{1}{2}\left(a_{12}^2 + a_{13}^2\right)^2, \quad D_d = a_{11}^4,$$

$$D_g = a_{11}^2\left(a_{12}^2 + a_{13}^2\right), \quad D_e = \tfrac{1}{2}\left(a_{12}^2 + a_{13}^2\right)^2,$$

$$D_h = D_g, \quad D_f = 2a_{11}^2\left(a_{12}^2 + a_{13}^2\right),$$

$$E_c = \left(a_{22}a_{32} + a_{23}a_{33}\right)^2, \quad E_d = 2a_{21}^2 a_{31}^2,$$

$$E_e = 2a_{21}a_{31}\left(a_{22}a_{32} + a_{23}a_{33}\right)$$

3.4 Optimization Problem for Variational Bounds

Variational bounds are derived based on the most rigorous theoretical foundations and usually serve as checkpoints for all other direct methods. Despite a significant amount of literature accumulated in the area of variational bounds (see, e.g., [30]), construction of bounds for the composite materials having prescribed microstructures has not been accomplished yet. In this work an attempt is made to derive such bounds by merging micromechanics and optimization (i.e. semidefinite programming) methodologies.

The theory of semidefinite programming (SDP) concerns solving optimization problems where the optimization variables are symmetric positive semidefinite matrices [10, 13]. Therefore, the optimization problem for determining bounds on the overall elastic moduli of a two-phase composite material can be formulated in the form:

$$\text{Upper bounds:} \quad \min_{L_0} \{C \bullet \bar{L}(L_0) \mid L_0 - L_i \succeq 0,\ i = 1, 2\} \tag{3.5}$$

$$\text{Lower bounds:} \quad \max_{L_0} \{C \bullet \bar{L}(L_0) \mid L_i - L_0 \succeq 0,\ i = 1, 2\} \tag{3.6}$$

where

$$\bar{L}(L_0) = \Big(f_1 \big\{ (L_0^* + L_1)^{-1} \big\} + f_2 (L_0^* + L_2)^{-1} \Big)^{-1} - L_0^*$$

is defined by Walpole (1969), matrix C is defined as $C = \mathbf{1}\mathbf{1}^T$, and $\bullet$ is the Frobenius inner product,

$$A \bullet B = \text{Tr}\, A^T B = \sum_i \sum_j A_{ij} B_{ij}$$

and $f_{1,2} \in [0, 1]$, $f_1 + f_2 = 1$ are given constants. Expression $\{A\}$ denotes *averaging* of tensor A and will be explained later in the paper.

Positive definite matrices L_1 and L_2 represent tensors of elastic moduli of the fibers

and matrix. We consider transversely isotrropic fibers dispersed in an isotropic matrix:

$$L_1 = \begin{bmatrix} n_1 & l_1 & l_1 & 0 & 0 & 0 \\ l_1 & m_1+k_1 & k_1-m_1 & 0 & 0 & 0 \\ l_1 & k_1-m_1 & m_1+k_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2m_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2p_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2p_1 \end{bmatrix} \tag{3.7}$$

$$L_2 = \begin{bmatrix} n_2 & l_2 & l_2 & 0 & 0 & 0 \\ l_2 & n_2 & l_2 & 0 & 0 & 0 \\ l_2 & l_2 & n_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & n_2-l_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & n_2-l_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & n_2-l_2 \end{bmatrix} \tag{3.8}$$

L_0 is the tensor of the form

$$L_0 = \begin{bmatrix} n_0 & l_0 & l_0 & 0 & 0 & 0 \\ l_0 & n_0 & l_0 & 0 & 0 & 0 \\ l_0 & l_0 & n_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & n_0-l_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & n_0-l_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & n_0-l_0 \end{bmatrix} \tag{3.9}$$

Tensor L_0^* has the same form as L_0, with n_0, l_0 replaced by n_0^*, l_0^*, respectively, where

$$l_0^* = \frac{2}{3}(n_0 - l_0) - \left(\frac{2}{n_0+l_0} + \frac{10}{7n_0+2l_0}\right)^{-1}$$
$$n_0^* = \frac{2}{3}(n_0 - l_0) + 2\left(\frac{2}{n_0+l_0} + \frac{10}{7n_0+2l_0}\right)^{-1}$$

3.5 Solving Optimization Problem

We consider a particular case of the optimization problem for finding the upper bound for the case of random distribution of fibers. In this case the optimization problem

(3.5) can be rewritten in the form:

$$\begin{aligned}
\min \quad & f(x) = C \bullet \Big(f_1\big\{(L_0^* + L_1)^{-1}\big\} + f_2(L_0^* + L_2)^{-1} \Big)^{-1} - L_0^* \\
\text{s.t.} \quad & 2l_0 - 2l_2 + n_0 - n_2 \geq 0, \\
& -l_0 + l_2 + n_0 - n_2 \geq 0, \\
& -l_0 - 2m_1 + n_0 \geq 0, \\
& -l_0 + n_0 - 2p_1 \geq 0, \\
& -2k_1 + l_0 + 2n_0 - n_1 - \big(4k_1^2 - 4k_1 l_0 + 9l_0^2 \\
& \qquad - 16l_0 l_1 + 8l_1^2 - 4k_1 n_1 + l_0 n_1 + n_1^2\big)^{1/2} \geq 0, \\
& -2k_1 + l_0 + 2n_0 - n_1 + \big(4k_1^2 - 4k_1 l_0 + 9l_0^2 \\
& \qquad - 16l_0 l_1 + 8l_1^2 - 4k_1 n_1 + l_0 n_1 + n_1^2\big)^{1/2} \geq 0,
\end{aligned} \tag{3.10}$$

where the semidefiniteness constraints are expressed explicitly using the elements of matrices L_0^*, L_1, and L_2. To solve this nonlinear SDP optimization problem, we employ an interior point method developed by Benson etal [9], which we briefly describe below.

Let us consider a nonlinear optimization problem

$$\begin{aligned}
\min \quad & f(\mathbf{x}) \\
\text{s.t.} \quad & \mathbf{h}(\mathbf{x}) \geq \mathbf{0}.
\end{aligned} \tag{3.11}$$

For this problem one can write the logarithmic barrier problem

$$\begin{aligned}
\min \quad & f(\mathbf{x}) - \mu \sum_{i=1}^{m} \log w_i \\
\text{s.t.} \quad & \mathbf{h}(\mathbf{x}) - \mathbf{w} = \mathbf{0}
\end{aligned} \tag{3.12}$$

and the Lagrangian function in the form

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{w}) = f(\mathbf{x}) - \mu \sum_{i=1}^{m} \log w_i - \mathbf{y}^\top (\mathbf{h}(\mathbf{x}) - \mathbf{w}),$$

where $\mathbf{y}$ are dual variables of (3.11). Next, from the first-order optimality conditions one can derive a primal-dual system

$$\nabla f(\mathbf{x}) - \nabla \mathbf{h}^\top \mathbf{y} = \mathbf{0},$$

$$-\mu \mathbf{e} + WY\mathbf{e} = \mathbf{0},$$

$$\mathbf{h}(\mathbf{x}) - \mathbf{w} = \mathbf{0}.$$

where $\nabla \mathbf{h}$ is the transpose of the Jacobian of the left-hand side of constraints (3.11); W and Y are diagonal matrices with the entries of $\mathbf{w}$ and $\mathbf{y}$, respectively: $W = \text{Diag}(\mathbf{w})$, $Y = \text{Diag}(\mathbf{y})$; and $\mathbf{e}$ is the vector of ones of an appropriate dimension.

This system can be solved to determine the optimal descent directions for $\mathbf{x}, \mathbf{y}$, and $\mathbf{w}$:

$$\Delta \mathbf{x} = (-H - \nabla \mathbf{h}^\top W^{-1} Y \nabla \mathbf{h})^{-1} (\boldsymbol{\sigma} - \nabla \mathbf{h}^\top W^{-1} Y \boldsymbol{\rho}),$$

$$\Delta \mathbf{y} = W^{-1} Y + \boldsymbol{\rho} + \boldsymbol{\gamma} - W^{-1} Y \nabla \mathbf{h}\, \Delta \mathbf{x},$$

$$\Delta \mathbf{w} = WY^{-1}(\boldsymbol{\gamma} - \Delta \mathbf{y}),$$

where

$$\boldsymbol{\sigma} = \nabla f(\mathbf{x}) - \nabla \mathbf{h}^\top \mathbf{y},$$

$$\boldsymbol{\gamma} = \mu W^{-1} \mathbf{e} - Y\mathbf{e},$$

$$\boldsymbol{\rho} = -\mathbf{h}(\mathbf{x}) + \mathbf{w},$$

$$H = \nabla^2 f(\mathbf{x}) - \sum_{i=1}^{m} y_i \nabla^2 h_i(\mathbf{x}).$$

Then, the values of optimization variables $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{w}$ are iteratively updated as:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \Delta \mathbf{x}^k,$$

$$\mathbf{y}^{k+1} = \mathbf{y}^k + \alpha^k \Delta \mathbf{y}^k,$$

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \alpha^k \Delta \mathbf{w}^k,$$

where $\alpha^k \in (0, 1)$ is the step size.

The described interior point algorithm is using first- and second-order information, e.g., $\nabla f(\mathbf{x})$, in order to compute the iterates of the solution. In this study, the corresponding derivatives were obtained numerically using the simple numerical differentiation scheme:

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h}$$

3.6 Hashin-Shtrikman-Walpole (HSW) Bounds

In the case of randomly distributed (three-dimensional, spatially uniform), transversely isotropic inclusions embedded in an isotropic matrix, there exist bounds derived by Walpole [41] and are often referred as the Hashin-Shtrikman-Walpole (HSW) bounds. These bounds are not optimal but rather feasible solutions of the corresponding optimization problems. Below we briefly discuss the HSW bounds.

Consider transversely isotropic inclusions randomly dispersed in an isotropic matrix. Denote the overall elastic moduli tensor of the resulting macroscopically isotropic

composite as

$$
\begin{aligned}
L &= \begin{bmatrix} n & l & l & 0 & 0 & 0 \\ l & m+k & k-m & 0 & 0 & 0 \\ l & k-m & m+k & 0 & 0 & 0 \\ 0 & 0 & 0 & 2m & 0 & 0 \\ 0 & 0 & 0 & 0 & 2p & 0 \\ 0 & 0 & 0 & 0 & 0 & 2p \end{bmatrix} \\
&= \begin{bmatrix} n & l & l & 0 & 0 & 0 \\ l & n & l & 0 & 0 & 0 \\ l & l & n & 0 & 0 & 0 \\ 0 & 0 & 0 & n-l & 0 & 0 \\ 0 & 0 & 0 & 0 & n-l & 0 \\ 0 & 0 & 0 & 0 & 0 & n-l \end{bmatrix}
\end{aligned}
\tag{3.13}
$$

Assume that transversely isotropic inclusions form the first phase, for which the elastic moduli tensor can be expressed by (3.7) and an isotropic matrix is the second phase, for which the elastic moduli tensor is can be expressed by (3.8).

Composite's bulk modulus, K, and shear modulus, G, are

$$
\begin{aligned}
K &= \frac{1}{9}(4k + 4l + n), \\
G &= \frac{1}{3}(k - 2l + n).
\end{aligned}
\tag{3.14}
$$

The HSW bounds of the composite are defined as

$$
\begin{aligned}
\bar{K} &= \left(\frac{f_1}{K^* + K_1 - a_1^2/(G^* + G_1)} + \frac{f_2}{K^* + K_2} \right)^{-1} - K^*, \\
\bar{G} &= \left[\frac{1}{5} f_1 \left(\frac{1}{G^* + G_1 - a_1^2/(K^* + K_1)} + \frac{2}{G^* + m_1} \right. \right. \\
&\quad \left. \left. + \frac{2}{G^* + p_1} \right) + \frac{f_2}{G^* + G_2} \right]^{-1} - G^*.
\end{aligned}
\tag{3.15}
$$

Here f_1 and f_2 are volume fractions of the inclusions (phase 1) and matrix (phase 2),

correspondingly, and $f_1 + f_2 = 1$. Moreover,

$$\begin{aligned} K_1 &= \frac{1}{9}(4k_1 + 4l_1 + n_1), \\ G_1 &= \frac{1}{3}(k_1 - 2l_1 + n_1), \\ K_2 &= \frac{1}{9}(4k_2 + 4l_2 + n_2) = \frac{1}{3}(n_2 + 2l_2), \\ G_2 &= \frac{1}{3}(k_2 - 2l_2 + n_2) = \frac{1}{2}(n_2 - l_2), \\ a_1^2 &= (n_1 + l_1 - 2k_1)^2/27 \end{aligned} \tag{3.16}$$

and

$$\begin{aligned} K^* &= \frac{4}{3}G^a, \\ G^* &= \frac{3}{2}\left[1/G^a + 10/(9K^a + 8G^a)\right]^{-1}. \end{aligned} \tag{3.17}$$

Parameters K^a and G^a in (3.17) are arbitrary and were selected by Walpole in order to provide the tight restrictive bounds (3.15). So, explicit formulae for them depend on the elastic properties of composite's phases. A general procedure to determine K^a and G^a may be found in the original paper by Walpole [41]. In particular, formula (3.15) will provide upper bounds for bulk and shear moduli if K^a and G^a are taken as

$$K^a = K_\upsilon + \upsilon, \quad G^a = G_\upsilon + \upsilon, \tag{3.18}$$

where

$$\begin{aligned} K_\upsilon &= \max\{K_1, K_2\}, \\ G_\upsilon &= \max\{G_1, m_1, p_1, G_2\} \\ \upsilon &= \begin{cases} 0, & K_\upsilon > K_1 \text{ and } G_\upsilon > G_1 \\ |a_1|, & \text{otherwise} \end{cases} \end{aligned} \tag{3.19}$$

The lower bounds for bulk and shear moduli will be obtained from (3.15) if K^a and G^a are selected as

$$K^a = \max\{0, K_\lambda - \lambda\}, \quad G^a = \max\{0, G_\lambda - \lambda\}, \tag{3.20}$$

where

$$K_\lambda = \min\{K_1, K_2\}, \tag{3.21}$$

$$G_\lambda = \min\{G_1, m_1, p_1, G_2\} \tag{3.22}$$

$$\lambda = \begin{cases} 0, & K_\lambda < K_1 \text{ and } G_\lambda < G_1 \\ |a_1|, & \text{otherwise} \end{cases} \tag{3.23}$$

Note that when $K_\lambda < \lambda$, $G_\lambda < \lambda$ in the expressions above, the corresponding lower bounds for K and G coincide with Reuss lower bounds.

$$\begin{gathered} K^a = K_\upsilon + \upsilon_1, \quad G^a = G_\upsilon + \upsilon_2, \\ K_\upsilon = \max\{K_1, K_2\}, \quad G_\upsilon = \max\{G_1, m_1, p_1, G_2\}. \end{gathered} \tag{3.24}$$

The lower bounds for bulk and shear moduli will be obtained from (3.15) if K^a and G^a are selected as

$$\begin{gathered} K^a = K_\lambda - \lambda_1, \quad G^a = G_\lambda - \lambda_2, \\ K_\lambda = \min\{K_1, K_2\}, \quad G_\lambda = \min\{G_1, m_1, p_1, G_2\}. \end{gathered} \tag{3.25}$$

Here the non-negative parameters $\upsilon_{1,2}$, $\lambda_{1,2} \geq 0$ should be taken as small as possible without violating the following inequalities

$$\begin{gathered} (K_\upsilon - K_1 + \upsilon_1)(G_\upsilon - G_1 + \upsilon_2) \geq a_1^2, \\ (K_1 - K_\lambda + \lambda_1)(G_1 - G_\lambda + \lambda_2) \geq a_1^2. \end{gathered} \tag{3.26}$$

If $K_\upsilon > K_1$, $G_\upsilon > G_1$, then $\upsilon_1 = \upsilon_2 = 0$ and, if $K_1 > K_\lambda$, $G_1 > G_\lambda$, then $\lambda_1 = \lambda_2 = 0$. Otherwise, υ_1 and/or υ_2 (λ_1 and/or λ_2) must be strictly positive and inequalities (3.26) should be changed to equalities. In this case the proper choice for parameters $\upsilon_{1,2}$, $\lambda_{1,2}$ is $\upsilon_1 = \upsilon_2 = \lambda_1 = \lambda_2 = |a_1|$.

For completeness, we also list the Voigt, $K_V,\ G_V$, and Reuss, $K_R,\ G_R$, bounds for a macroscopically isotropic composite with randomly (three-dimensional, spatially uniform) distributed transversely isotropic inclusions are expressible as:

$$\begin{aligned} K_V &= f_1 K_1 + f_2 K_2, \\ G_V &= \frac{1}{5} f_1 (G_1 + 2m_1 + 2p_1) + f_2 G_2, \\ \frac{1}{K_R} &= f_1 \frac{3G_1}{k_1 E_1} + \frac{f_2}{K_2}, \\ \frac{1}{G_R} &= \frac{1}{5} f_1 \left(\frac{3K_1}{k_1 E_1} + \frac{2}{m_1} + \frac{2}{p_1} \right) + \frac{f_2}{G_2}, \end{aligned} \tag{3.27}$$

where $E_1 = n_1 - l_1^2/k_1$, K_1 and G_1 are defined by (3.16). The HSW bounds are tighter than the Voight-Reuss bounds.

It is important to emphasize again that the HSW bounds developed by Walpole and summarized in this section represent *feasible*, but not necessarily *optimal* solutions to the general nonlinear SDP formulations (3.5)–(3.6). Moreover, the procedure developed by Walpole yields explicit bounds only in the case of uniformly distributed inclusions and is not suitable to derive bounds for arbitrary ODFs. In contrast, our semidefinite programming formulation (3.5) and (3.6) enables to construct bounds for composites with arbitrary distributed phases.

3.7 Analysis of SDP bounds

Alternative formulation of optimization problem (3.5) for upper bounds on the overall elastic moduli K, G:

$$\begin{aligned}
\min_{G^a,K^a} \quad & 3\bar{K} + 2\bar{G} \\
\text{s.t.} \quad & K^a - K_2 \geq 0 \\
& G^a - G_2 \geq 0 \\
& G^a - m_1 \geq 0 \\
& G^a - p_1 \geq 0 \\
& K^a + \tfrac{1}{3}G^a - k_1 \geq 0 \\
& K^a + \tfrac{4}{3}G^a - n_1 \geq 0 \\
& (K^a + \tfrac{1}{3}G^a - k_1)(K^a + \tfrac{4}{3}G^a - n_1) \geq (K^a - \tfrac{2}{3}G^a - l_1)^2
\end{aligned} \tag{3.28}$$

where $\bar{K}$, $\bar{G}$ are given by (3.15) and (3.17), K_i, G_i are related to the components of tensors L_1, L_2 by (3.16), and the optimization variables K^a, G^a are related to the components of tensor L_0 in (3.5) as

$$K^a = K_0 = \frac{n_0 + 2l_0}{3}, \quad G^a = G_0 = \frac{n_0 - l_0}{2}$$

It is easy to see that the corresponding formulation (3.6) of the lower bound SDP

problem reduces to a nonlinear programming problem for lower bounds

$$\begin{aligned}
\max_{G^a,K^a} \quad & 3\bar{K} + 2\bar{G} \\
\text{s.t.} \quad & K^a - K_2 \leq 0 \\
& G^a - G_2 \leq 0 \\
& G^a - m_1 \leq 0 \\
& G^a - p_1 \leq 0 \\
& K^a + \tfrac{1}{3}G^a - k_1 \leq 0 \\
& K^a + \tfrac{4}{3}G^a - n_1 \leq 0 \\
& (K^a + \tfrac{1}{3}G^a - k_1)(K^a + \tfrac{4}{3}G^a - n_1) \geq (K^a - \tfrac{2}{3}G^a - l_1)^2
\end{aligned} \tag{3.29}$$

Note that the second order cone constraint in both upper and lower bound formulations (3.28), (3.29) is unchanged. Clearly, both these problems have feasible regions that are convex in K^a, G^a, and a common nonlinear objective function

$$F(K^a, G^a) = 3\bar{K}(K^a, G^a) + 2\bar{G}(K^a, G^a) \tag{3.30}$$

where $\bar{K}$ and $\bar{G}$ are given by (3.15) and (3.17).

The optimal solutions of problems (3.28), (3.29) can be derived through the analysis of the objective function and feasible regions of the respective problems. First, we show that the highly nonlinear objective function F has a rather simple structure.

Proposition 5. *Functions $\bar{K}$ and $\bar{G}$* (3.15)*, and consequently the objective function F in problems* (3.28)*,* (3.29) *are non-decreasing in both K^a and G^a.*

Proof. From (3.17) it is easy to see that

$$\begin{aligned}
&\frac{\partial K^*}{\partial K^a} = 0,\\
&\frac{\partial K^*}{\partial G^a} = \frac{4}{3} > 0,\\
&\frac{\partial G^*}{\partial K^a}\frac{3}{2}\left[1/G^a + 10/(9K^a + 8G^a)\right]^{-2}\frac{10 * 9}{(9K^a + 8G^a)^2} > 0,\\
&\frac{\partial G^*}{\partial G^a}\frac{3}{2}\left[1/G^a + 10/(9K^a + 8G^a)\right]^{-2}\frac{1}{(G^a)^2} > 0.
\end{aligned}$$

Next we show that partial derivatives of $\bar{K}$ and $\bar{G}$ with respect to K^* and G^* are non-negative. Consider, for example, $\partial\bar{G}/\partial G^*$. By introducing notation

$$\begin{aligned}
A_1 &= \frac{1}{G^* + G_1 - a_1^2/(K^* + K_1)},\\
A_2 &= \frac{1}{G^* + m_1}\\
A_3 &= \frac{1}{G^* + p_1}\\
A_4 &= \frac{1}{G^* + G_2}
\end{aligned}$$

and noting that $\dfrac{\partial A_i}{\partial G^*} = -A_i^2$, one has

$$\begin{aligned}
\frac{\partial\bar{G}}{\partial G^*} &= -\frac{-\frac{1}{5}f_1A_1^2 - \frac{2}{5}f_1A_2^2 - \frac{2}{5}f_1A_3^2 - f_2A_4^2}{\left(\frac{1}{5}f_1A_1 + \frac{2}{5}f_1A_2 + \frac{2}{5}f_1A_3 + f_2A_4\right)^2} - 1\\
&= \left(\frac{1}{5}f_1A_1 + \frac{2}{5}f_1A_2 + \frac{2}{5}f_1A_3 + f_2A_4\right)^{-2}V
\end{aligned}$$

where expression V is given by

$$V = \left(\frac{1}{5}f_1A_1^2 + \frac{2}{5}f_1A_2^2 + \frac{2}{5}f_1A_3^2 + f_2A_4^2\right) - \left(\frac{1}{5}f_1A_1 + \frac{2}{5}f_1A_2 + \frac{2}{5}f_1A_3 + f_2A_4\right)^2$$

It is easy to see that $V \geq 0$ in general, and, moreover, $V = 0$ if and only if $A_1 = \ldots = A_4$. Namely, V can be viewed as the variance of a random variable A that takes values A_1, A_2,

A_3, A_4 with probabilities $\frac{1}{5}f_1$, $\frac{2}{5}f_1$, $\frac{2}{5}f_1$, f_2, respectively:

$$V = \mathsf{Var}(A) = \mathsf{E}(A^2) - (\mathsf{E}A)^2 \geq 0.$$

Similarly, it can be shown that $\partial\bar{K}/\partial K^*$ is nonnegative.

To derive expression for $\partial\bar{G}/\partial K^*$, denote

$$[\ldots] = \left[\frac{1}{5}f_1\left(\frac{1}{G^* + G_1 - a_1^2/\left(K^* + K_1\right)} + \frac{2}{G^* + m_1} + \frac{2}{G^* + p_1}\right) + \frac{f_2}{G^* + G_2}\right],$$

then

$$\begin{aligned}
\frac{\partial\bar{G}}{\partial K^*} &= (-1)\left[\ldots\right]^{-2}\frac{\partial}{\partial K^*}\left[\ldots\right] = \\
&= (-1)\left[\ldots\right]^{-2}\left[\frac{f_1}{5}(-1)(-a_1^2)(-1)\right]\frac{1}{\left(G^* + G_1 - a_1^2/\left(K^* + K_1\right)\right)^2}\frac{1}{\left(K^* + K_1\right)^2} = \\
&= \left[\ldots\right]^{-2}\frac{1}{5}f_1 a_1^2\frac{1}{\left(G^* + G_1 - a_1^2/\left(K^* + K_1\right)\right)^2}\frac{1}{\left(K^* + K_1\right)^2} \geq 0.
\end{aligned}$$

Nonnegativity of $\partial\bar{K}/\partial G^*$ can be proved the same way.

Note that it is possible that partial derivatives equal zero. $\partial\bar{K}/\partial G^*$ and $\partial\bar{G}/\partial K^*$ equal zero if and only if $f_1 = 0$. Partial derivative $\partial\bar{K}/\partial K^*$ equals zero if and only if

$$\frac{1}{K^* + K_1 - a_1^2/(G^* + G_1)} = \frac{1}{K^* + K_2}$$

$$G^* = a_1^2/(K_1 - K_2) - G_1$$

Partial derivative $\partial\bar{G}/\partial G^*$ equals zero if and only if

$$\frac{1}{G^* + G_1 - a_1^2/(K^* + K_1)} = \frac{1}{G^* + m_1} = \frac{1}{G^* + p_1} = \frac{1}{G^* + G_2}$$

$$G^* + G_1 - a_1^2/(K^* + K_1) = G^* + m_1 = G^* + p_1 = G^* + G_2$$

$$m_1 = p_1 = G_2 = G_1 - a_1^2/(K^* + K_1)$$

$$m_1 = p_1 = G_2 \quad K^* = a_1^2/(G_1 - G_2) - K_1$$

In conclusion, all partial derivatives are nonnegative, whence $\bar{K}$, $\bar{G}$ are non-decreasing in K^a, G^a.

Proposition 6. *The (global) optimal solutions of upper and lower bound problems* (3.28)*,* (3.29) *are located at the boundaries of their respective feasible regions.*

Proof. As shown in 5, $\bar{K}$, $\bar{G}$ are non-decreasing in K^a, G^a. The objective function for (3.29) and (3.28) is a linear combination of $\bar{K}$, $\bar{G}$ and hence it is monotonically non-decreasing in K^a and G^a. Therefore, optimum point is located at the boundaries of feasible region.

3.7.1 Analysis of feasible region

As it has been shown in Proposition 6, optimal solutions of problems (3.28) and (3.29) are located at the boundaries of the respective feasible sets. To facilitate analysis of the feasible sets of problems (3.28)–(3.29), denote

$$K^a = K_1 + Y_1, \quad G^a = G_1 + Y_2. \tag{3.31}$$

Then, elementary algebraic manipulations yield that

$$\begin{aligned} K^a + \frac{1}{3}G^a - k_1 &= Y_1 + \frac{1}{3}Y_2 - 2C \\ K^a + \frac{4}{3}G^a - n_1 &= Y_1 + \frac{4}{3}Y_2 + 4C \\ K^a - \frac{2}{3}G^a - l_1 &= Y_1 - \frac{2}{3}Y_2 + C, \end{aligned}$$

where it is denoted

$$C = \frac{1}{9}(2k_1 - l_1 - n_1). \tag{3.32}$$

Consequently, the last three constraints of problem (3.28) (respectively, (3.29)) take the form

$$Y_1 + \frac{1}{3}Y_2 - 2C \geq (\leq)\ 0 \tag{3.33}$$

$$Y_1 + \frac{4}{3}Y_2 + 4C \geq (\leq)\ 0 \tag{3.34}$$

$$Y_1 Y_2 \geq 3C^2 \tag{3.35}$$

It is easy to see that lines $Y_1 + \frac{1}{3}Y_2 - 2C = 0$ and $Y_1 + \frac{4}{3}Y_2 + 4C = 0$ represent supporting hyperplanes of surfaces $Y_1Y_2 - 3C^2 = 0$ at points $(C, 3C)$ and $(-2C, -\frac{3}{2}C)$, respectively. Particularly, if $C > 0$, then the line $Y_1 + \frac{1}{3}Y_2 - 2C = 0$ is tangent to the hyperbola $\{Y_1Y_2 = 3C^2;\ Y_1, Y_2 \geq 0\}$ at the point $(C, 3C)$, while the line $Y_1 + \frac{4}{3}Y_2 + 4C = 0$ is tangent to the hyperbola $\{Y_1Y_2 = 3C^2;\ Y_1, Y_2 \leq 0\}$ at the point $(-2C, -\frac{3}{2}C)$.

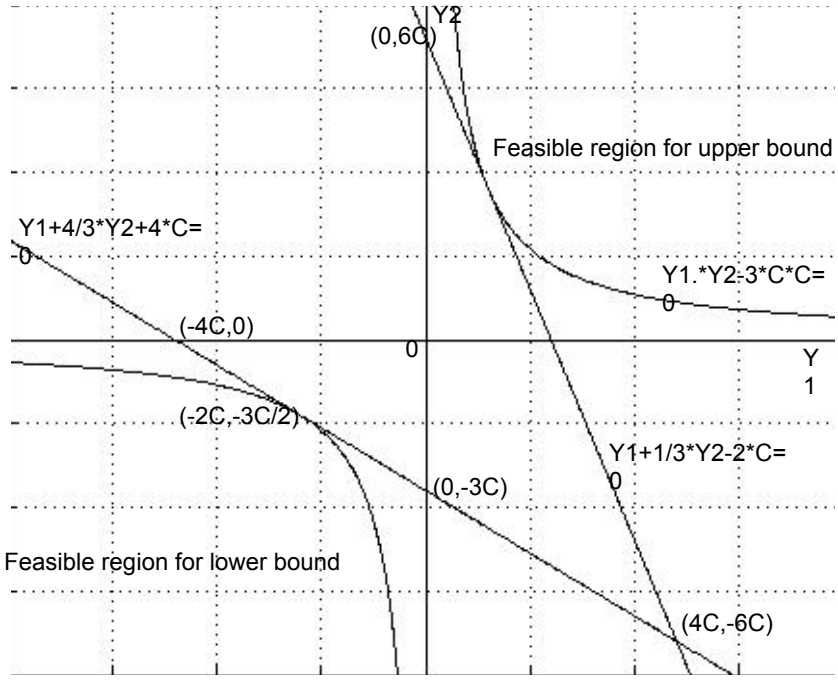
Similarly, if $C < 0$, then the line $Y_1 + \frac{1}{3}Y_2 - 2C = 0$ is tangent to the hyperbola $\{Y_1Y_2 = 3C^2;\ Y_1, Y_2 \leq 0\}$ at the point $(C, 3C)$, and the line $Y_1 + \frac{4}{3}Y_2 + 4C = 0$ is tangent to the hyperbola $\{Y_1Y_2 = 3C^2;\ Y_1, Y_2 \geq 0\}$ at the point $(-2C, -\frac{3}{2}C)$.

In these two cases, constraints (3.33) define region $\{Y_1Y_2 \geq 3C^2;\ Y_1, Y_2 \geq 0\}$ in the upper bound problem (3.28), and $\{Y_1Y_2 \geq 3C^2;\ Y_1, Y_2 \leq 0\}$ in the lower bound problem (3.29).

Finally, if $C = 0$, then inequalities (3.33) correspond to $\{Y_1 \geq 0,\ Y_2 \geq 0\}$ for the upper bound problem (3.28), and $\{Y_1 \leq 0,\ Y_2 \leq 0\}$ for (3.29).

The feasible regions of problems (3.28) and (3.29), respectively, can be expressed

Figure 3.4: Feasible region for upper and lower bounds for $C > 0$



as

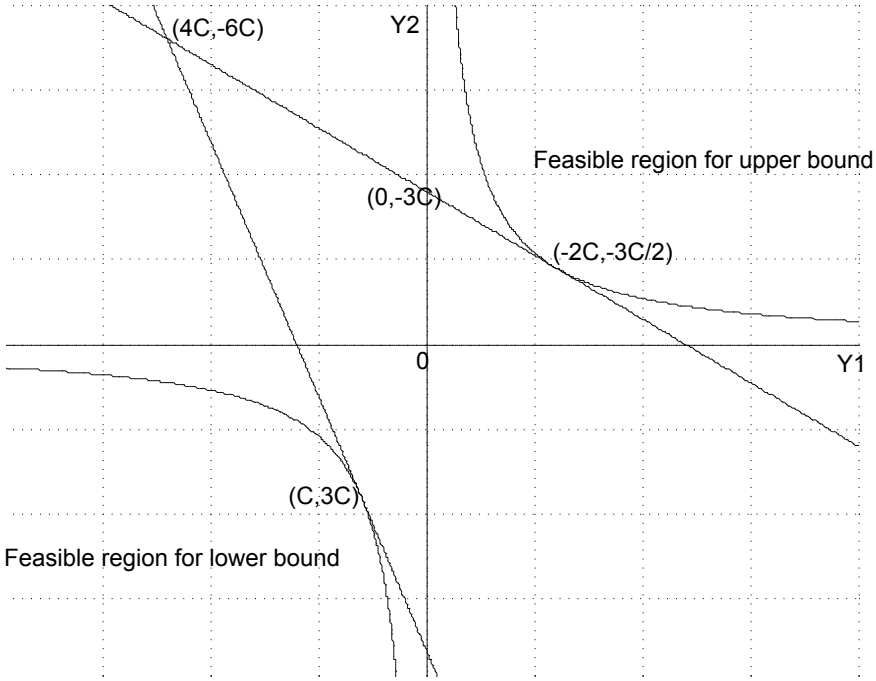
$$Y_1 \geq K_2 - K_1 \tag{3.36}$$

$$Y_2 \geq \max\{G_2,\, m_1,\, p_1\} - G_1 \tag{3.37}$$

$$Y_1 Y_2 \geq a_1^2 \tag{3.38}$$

$$Y_1 \geq 0,\, Y_2 \geq 0 \tag{3.39}$$

Figure 3.5: Feasible region for upper and lower bounds for $C < 0$



and

$$Y_1 \leq K_2 - K_1 \tag{3.40}$$

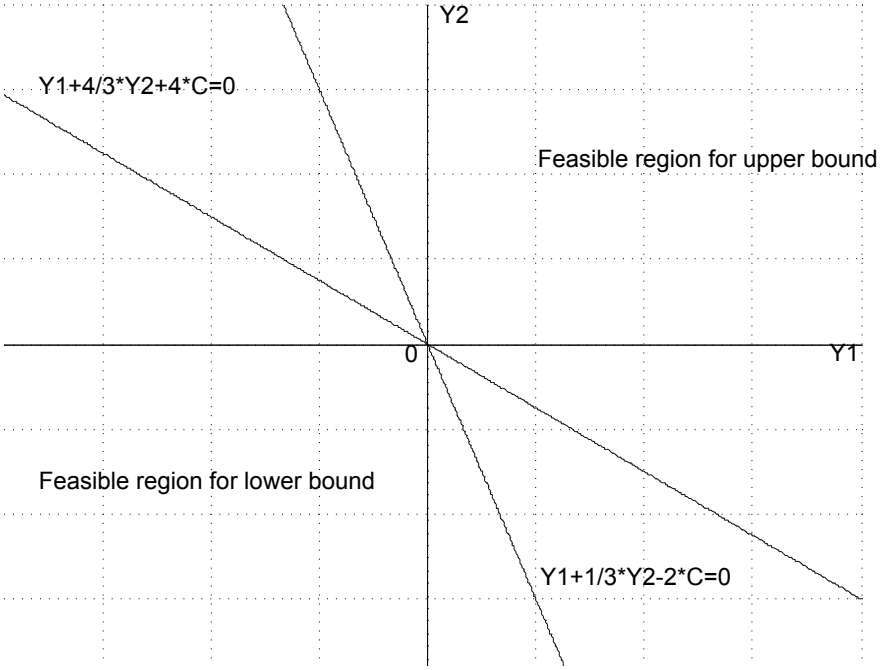
$$Y_2 \leq \min\{G_2,\, m_1,\, p_1\} - G_1 \tag{3.41}$$

$$Y_1 Y_2 \geq a_1^2 \tag{3.42}$$

$$Y_1 \leq 0,\; Y_2 \leq 0 \tag{3.43}$$

For further analysis it is convenient to rewrite the above representation of the feasi-

Figure 3.6: Feasible region for upper and lower bounds for $C = 0$



ble region of problem (3.28) in the form

$$\begin{aligned} Y_1 &\geq \max\{0, K_2 - K_1\} = \max\{K_1, K_2\} - K_1 \\ Y_2 &\geq \max\{0, \max\{G_2,\, m_1,\, p_1\} - G_1\} = \max\{G_1, G_2, m_1, p_1\} - G_1 \\ Y_1 Y_2 &\geq a_1^2 \end{aligned} \tag{3.44}$$

3.7.1.0.1 Analysis of solution to upper bound problem (3.28)

In what follows, we consider the following cases.

Case U1: Observe that due to Proposition 6, the optimal solution of (3.28) is given by

$$Y_1^* = K_2 - K_1, \quad Y_2^* = \max\{G_2, m_1, p_1\} - G_1, \tag{3.45}$$

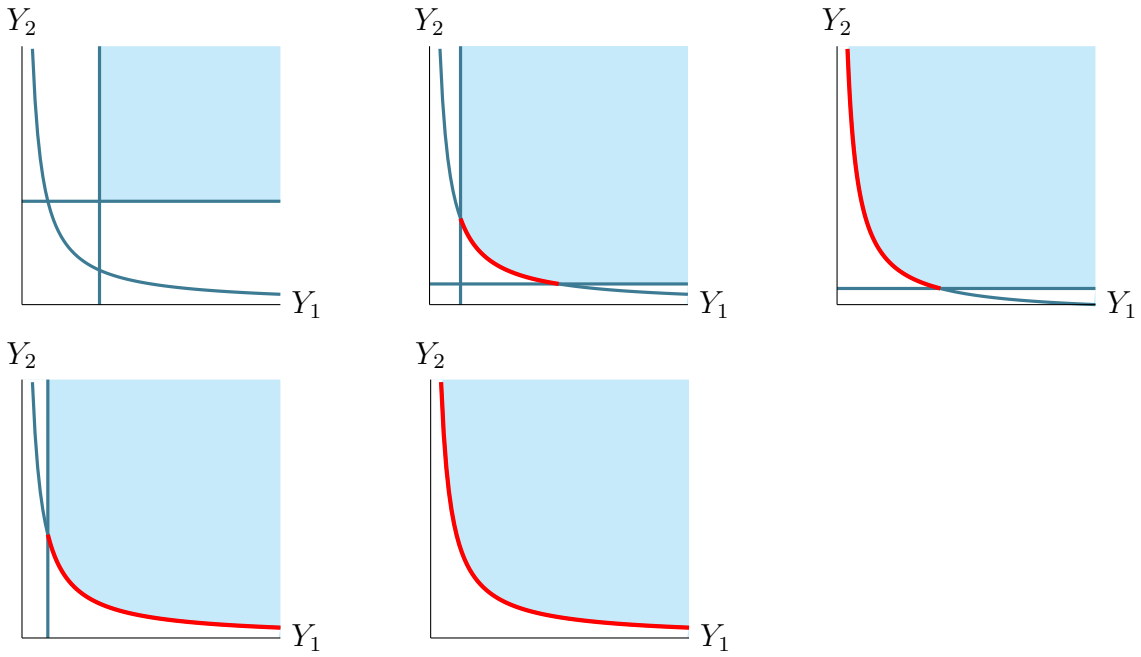
if the following conditions are satisfied:

$$\begin{cases} K_2 > K_1, \\ \max\{G_2, m_1, p_1\} > G_1, \\ (K_2 - K_1)(\max\{G_2, m_1, p_1\} - G_1) \geq a_1^2. \end{cases} \tag{3.46}$$

It is easy to see that the optimal solution (3.45)–(3.46) coincides with Walpole [41] solution (3.18)–(3.19) corresponding to $v = 0$ in (3.18):

$$\begin{aligned} K^a = K_2 \Leftrightarrow Y_1^W = K_2 - K_1, \\ G^a = \max\{G_2, m_1, p_1\} \Leftrightarrow Y_2^W = \max\{G_2, m_1, p_1\} - G_1 \end{aligned} \tag{3.47}$$

Figure 3.7: Feasible region and optimal solution regions in cases U1-U5



Note, however, that the third condition in (3.46) is disregarded in Walpole's expressions (3.18)–(3.19). If this condition is violated, it is possible that Walpole's solution (3.47)

is infeasible to the upper bound problem (3.28), as is shown next.

Case U2: If, on the other hand, one has that

$$\begin{cases} K_2 > K_1, \\ \max\{G_2, m_1, p_1\} > G_1, \\ (K_2 - K_1)(\max\{G_2, m_1, p_1\} - G_1) < a_1^2, \end{cases} \tag{3.48}$$

then an optimal solution of the upper bound problem (3.28) is located on a segment of the hyperbola $Y_1 Y_2 = a_1^2$:

$$\left\{Y_1 \geq 0, Y_2 \geq 0 : Y_1 Y_2 = a_1^2,\ Y_1 \geq K_2 - K_1,\ Y_2 \geq \max\{G_2, m_1, p_1\} - G_1\right\} \tag{3.49}$$

In this case, Walpole's procedure (3.18)–(3.19) yields the same values values (3.47) that are infeasible for the upper bound problem (3.28) due to violation of the conic constraint, and thus lead to a tensor of material properties that is not guaranteed to be positive definite.

Next we show that the set of material properties $\{k_i, l_i, n_i, m_i, p_i\}$, $i = 1, 2$, that satisfy conditions (3.48) is non-empty. Indeed, assume that material of inclusions w(material 1) is such that

$$\max\{m_1, p_1\} \leq G_1 = \frac{1}{3}(k_1 - 2l_1 + n_1).$$

Note that an isotropic material satisfies the above condition and consider a base material (material 2) with parameters K_2, G_2 such that

$$K_2 = (1 + \varepsilon)K_1, \quad G_2 = (1 + \delta)G_1$$

for some specific $\varepsilon > 0$ and $\delta > 0$. Obviously, one has that $\max\{G_2, m_1, p_1\} = G_2$ due to the above assumption on G_1. Then, one immediately has that the difference

$$(K_2 - K_1)(\max\{G_2, m_1, p_1\} - G_1) - a_1^2 = \varepsilon\delta K_1 G_1 - a_1^2$$

$$= \varepsilon\delta\frac{1}{27}(4k_1 + 4l_1 + n_1)(k_1 - 2l_1 + n_1) - \frac{1}{27}(n_1 + l_1 - 2k_1)^2,$$

can be made negative for any given set of parameters of material 1 by choosing sufficiently small $\varepsilon, \delta > 0$.

Case U3:

$$\begin{cases} K_2 \leq K_1, \\ \max\{G_2, m_1, p_1\} > G_1, \end{cases} \tag{3.50}$$

whereby the optimal solution of (3.28) lies on the segment of the hyperbola:

$$\left\{Y_1, Y_2 \geq 0 : Y_1 Y_2 = a_1^2, Y_2 \geq \max\{G_2, m_1, p_1\} - G_1\right\}$$

Walpole solution is always feasible and non-optimal:

$$K^a = \max\{K_1, K_2\} + |a_1| = K_1 + |a_1| \Leftrightarrow Y_1^W = |a_1|,$$

$$G^a = \max\{G_1, G_2, m_1, p_1\} + |a_1| \Leftrightarrow Y_2^W = \max\{G_2, m_1, p_1\} - G_1 + |a_1|$$

Case U4:

$$\begin{cases} K_2 > K_1, \\ \max\{G_2, m_1, p_1\} \leq G_1, \end{cases} \tag{3.51}$$

whereby the optimal solution of (3.28) lies on the segment of the hyperbola:

$$\left\{Y_1, Y_2 \geq 0 : Y_1 Y_2 = a_1^2, Y_1 \geq K_2 - K_1\right\}$$

Walpole solution is always feasible and non-optimal:

$$K^a = \max\{K_1, K_2\} + |a_1| = K_2 + |a_1| \Leftrightarrow Y_1^W = K_2 - K_1 + |a_1|,$$

$$G^a = \max\{G_1, G_2, m_1, p_1\} + |a_1| = G_1 + |a_1| \Leftrightarrow Y_2^W = |a_1|$$

Case U5:

$$\begin{cases} K_2 \leq K_1, \\ \max\{G_2, m_1, p_1\} \leq G_1, \end{cases} \tag{3.52}$$

In this case, an optimal solution of upper bound problem (3.28) lies on the hyperbola

$$\{Y_1, Y_2 \geq 0 : Y_1 Y_2 = a_1^2\},$$

whereas Walpole's solution (3.18)–(3.19) represents the point $(Y_1^W, Y_2^W) = (|a_1|, |a_1|)$ of this hyperbola:

$$K^a = \max\{K_1, K_2\} + |a_1| = K_1 + |a_1| \Leftrightarrow Y_1^W = |a_1|,$$

$$G^a = \max\{G_1, G_2, m_1, p_1\} + |a_1| = G_1 + |a_1| \Leftrightarrow Y_2^W = |a_1|$$

3.8 Computational Results

Problems (3.28) and (3.29) were solved using nonlinear solver, computational results show that the obtained solution of the nonlinear SDP problem (3.10) yields improved upper and lower bound on the effective elastic modules K and G, as compared to bounds due to Voigt and Walpole (Figures 3.8 and 3.9).

Figure 3.8: Upper and lower bounds on the overall bulk modulus K of a two-phase fiber reinforced composite

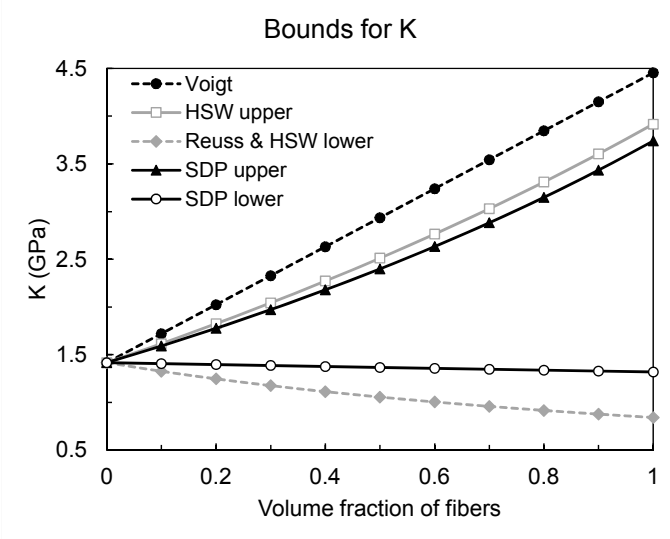
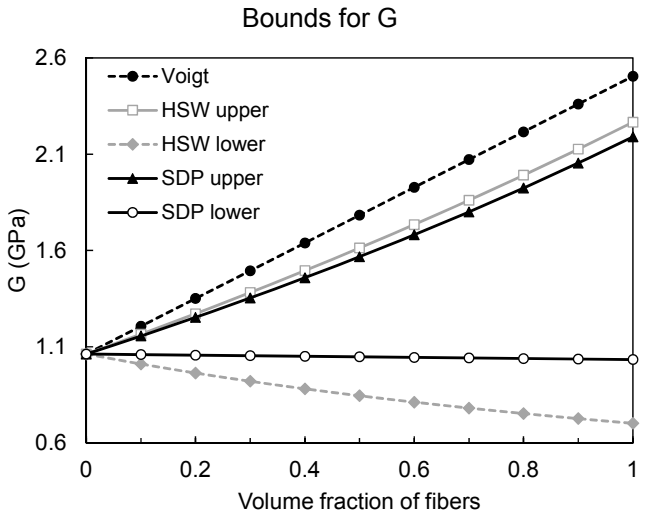


Figure 3.9: Upper and lower bounds on the overall bulk modulus G of a two-phase fiber reinforced composite



CHAPTER 4
STOCHASTIC ORDERINGS FOR MULTIOBJECTIVE OPTIMIZATION

4.1 Introduction

Design problems often require optimization of various parameters simultaneously. A multiobjective optimization problem in the most general form can be written as

$$\begin{aligned} \max \quad & \mathbf{f}(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{x} \in \mathcal{X} \end{aligned} \tag{4.1}$$

where $\mathcal{X} \subset \mathbb{R}^n$ is the set of feasible designs, and $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))^\top$ is a vector of objective functions, or design criteria to be maximized. Similarly, the vector-valued function $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_\ell(\mathbf{x}))$ represents the set of constraints on the design vector $\mathbf{x}$. Clearly, multiobjective optimization is usually associated with tradeoff between conflicting objectives, for instance minimizing cost while maximizing efficiency of the system. Particular design $\mathbf{x}^* \in \mathcal{X}$ is said to be efficient, or Pareto optimal if it satisfies the design constraints, $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, and there is no $\mathbf{x} \in \mathcal{X}$ that satisfies $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ and $\mathbf{f}(\mathbf{x}) \geq \mathbf{f}(\mathbf{x}^*)$, with at least one scalar inequality being strict in the latter vector inequality.

Numerical procedures for solving multi-objective optimization problems of type (4.1) typically rely on scalarization techniques, i.e., on reduction of (4.1) to a problem with a single objective. One of such scalarization techniques transforms (4.1) into a problem of

the form

$$\begin{aligned} \max \quad & \sum_{i=1}^{m} \mu_i f_i(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{x} \in \mathcal{X} \end{aligned} \tag{4.2}$$

where μ_i, $i = 1, \ldots, m$, are the "importance weights" of design criteria f_i. Another popular scalarization method consists in converting the objective functions into constraints:

$$\begin{aligned} \max \quad & f_k(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq \lambda_i, \quad i = 1, \ldots, m, \ i \neq k \\ & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{4.3}$$

for some fixed $k \in \{1, \ldots, m\}$, and preselected values λ_i, $i \neq k$.

By varying the parameters μ_i in (4.2) and λ_i in (4.3), one obtains a surface (manifold) in $\mathbb{R}^m$ that is the image of the set $\mathcal{X}^* \subset \mathcal{X} \subset \mathbb{R}^m$ of Pareto-optimal designs. Then, one may select a specific design $\mathbf{x}^*$ from the Pareto-optimal set $\mathcal{X}^*$ based on the desired combinations of achieved values of design criteria functions $f_i(\mathbf{x}^*)$.

In this Chapter we consider an alternative approach to scalarizing the multiobjective optimization problems of type (4.1) by exploring an analogy with *stochastic orderings* and related *utility theory* concepts. The motivation for the proposed approach comes from the recent interest in *multifunctional* structures and materials, which are capable of performing more than one function, for example, carrying load and storing energy. In this regard, it is of interest to determine whether the current design is "multifunctional", or to determine which of two multifunctional designs is "better", etc.

We expect that the developed approach will be especially useful in situations when improvement over a "benchmark" design or solution is not possible for every objective, thus requiring a sophisticated tradeoff principle for improving some of the objectives while allowing a carefully chosen deterioration in others.

4.2 Stochastic Dominance

Stochastic dominance concepts represent the way of ordering random variables that generalizes the standard ordering ">"on real line. The simplest case of stochastic dominance is first-order stochastic dominance (FSD): random variable X is said to dominate a random variable Y, with respect to FSD, $X \succeq_{(1)} Y$, if it generally assumes greater values, i.e.

$$\mathsf{P}(X \leq t) \leq \mathsf{P}(Y \leq t) \quad \text{for all} \quad t \in \mathbb{R},$$

or, using the cumulative distribution functions (CDF) $F_X(t)$ and $F_Y(t)$ of random variables X and Y, respectively,

$$X \succeq_{(1)} Y \quad \Leftrightarrow \quad F_X(t) \leq F_Y(t) \quad \text{for all} \quad t \in \mathbb{R}.$$

To illustrate the meaning of FSD, let t be an income level, then $X \succeq_{(1)} Y$ means that for any t the proportion of individuals with income greater than t in distribution X is higher or equal to the proportion of individuals with income greater than t from distribution Y. Or, if income smaller than t means poverty, the proportion of poor people in Y is higher or equal to the proportion of poor people in X. Essentially, for any poverty threshold, poverty level in Y is always higher than in X.

First order stochastic dominance establishes which random variable is "larger", and

the next step is to determine which of two random variables is "less risky", and here the second order stochastic dominance (SSD) comes into play. It is said that the random variable X dominates a random variable Y with respect to SSD, which is denoted by $X \succeq_{(2)} Y$, if

$$\int_{-\infty}^{t} F_X(\xi)d\xi \leq \int_{-\infty}^{t} F_Y(\xi)d\xi \quad \text{for all} \quad t \in \mathbb{R},$$

i.e., lower realizations of X are allowed with "low" probability.

In general, the *k-th order stochastic dominance (kSD)*, $k \geq 2$, has the form

$$X \succeq_{(k)} Y \quad \Leftrightarrow \quad F_X^{(k)}(t) \leq F_Y^{(k)}(t) \quad \text{for all} \quad t \in \mathbb{R}, \tag{4.4}$$

where function $F^{(k)}(t)$ is known as the k-th degree distribution function and is defined recursively as

$$F_X^{(k)}(t) = \int_{-\infty}^{t} F_X^{(k-1)}(\xi)\,\mathrm{d}\xi, \quad F_X^{(1)}(t) = F_X(t). \tag{4.5}$$

In practice, stochastic dominance is usually implemented through its connection to the expected utility theory of von Neumann and Morgenstern [40]. Let utility function $u(\cdot)$ be a representation of decision maker's preference for different values of the argument: $x \in \mathbb{R}$ is preferred over $y \in \mathbb{R}$ if $u(x) > u(y)$. The following Proposition shows the connection between the First Order Stochastic Dominance and utility theory:

Proposition 7. *$X \succeq_{(1)} Y$ if and only if for all non-decreasing utility functions u the following inequality holds:*

$$\mathsf{E}[u(X)] \geq \mathsf{E}[u(Y)]. \tag{4.6}$$

In [37] authors connected utility theory and SSD:

Proposition 8. *X dominates Y in SSD sense, $X \succeq_{(2)} Y$, if and only if* (4.6) *holds for all concave non-decresing functions u.*

In general, relation $X \succeq_{(k)} Y$ equivalent to the inequality (4.6) holding for all utility functions u from a certain class. As stated above, for FSD u should be nondecreasing, for SSD u should be nondecreasing and concave, for third order u is nondecresing concave and has positive third derivative, and so on. Thus to find non-dominated alternatives X^* from a given feasible set $\mathcal{X}$ one should solve the maximization problem:

$$\max_{X \in \mathcal{X}} E[u(X)] = \max_{X \in \mathcal{X}} \sum_{i=1}^{m} P(\omega_i) u(X(\omega_i))$$

with the appropriate utility function.

4.2.1 SD for multiobjective optimization

The stochastic dominance concepts can be employed to introduce ordering, or preference relations in multidimensional space $\mathbb{R}^m$. Namely, random variables can be regarded as points in $\mathbb{R}^m$ if the space of random events Ω is finite, $\Omega = \{\omega_1, \ldots, \omega_m\}$, and the associated probability measure P is fixed, $\mathsf{P}(\omega_i) = p_i > 0$, $i = 1, \ldots, m$, $p_1 + \ldots + p_m = 1$. In such a probability space, a random variable $Z : \Omega \mapsto \mathbb{R}$ has m realizations $z_i = Z(\omega_i)$, $i = 1, \ldots, m$, i.e., random variable Z can be represented by the vector of its realizations $(z_1, \ldots, z_m) = \mathbf{z} \in \mathbb{R}^m$.

In the present context, design $\mathbf{x}'$ is clearly preferred to design $\mathbf{x}''$ if the corresponding points in $\mathbb{R}^m$ defined by the m design criteria satisfy $\mathbf{f}(\mathbf{x}') \geq \mathbf{f}(\mathbf{x}'')$, i.e.,

$$f_i(\mathbf{x}') \geq f_i(\mathbf{x}''), \quad i = 1, \ldots, m.$$

In practice, however, design improvements often involve tradeoffs, hence a theoretically rigorous method of selecting the best design alternative is desirable in cases when

$$f_i(\mathbf{x}') > f_i(\mathbf{x}'') \text{ but } f_j(\mathbf{x}) < f_j(\mathbf{x}'') \text{ for some } i \text{ and } j.$$

In the present context, this translates into the possibility of comparing designs across multiple design criteria. For example, ideally one would prefer design $\mathbf{x}'$ to design $\mathbf{x}''$ if $f_i(\mathbf{x}') \geq f_i(\mathbf{x}'')$ for all criteria f_i, $i = 1, \ldots, m$. In practice, however, design improvements often involve tradeoffs, therefore "FSD-", "SSD-" and "kSD-based" rules can provide a theoretically sound way to select the overall better design in the cases when $f_i(\mathbf{x}') > f_i(\mathbf{x}'')$ but $f_j(\mathbf{x}) < f_j(\mathbf{x}'')$ for some i and j.

As an important consequence of the described connection between the utility theory and stochastic orderings, the set of "efficient", or *non-dominated* elements X from a given set of alternatives $\mathfrak{X}$ can be determined by solving the *expected utility maximization problem* with an appropriately chosen utility function u:

$$\max_{X \in \mathfrak{X}} \mathsf{E}[u(X)] = \max_{X \in \mathfrak{X}} \sum_{i=1}^{m} \mathsf{P}(\omega_i) u(X(\omega_i)). \tag{4.7}$$

To apply the above in the context of multiobjective optimization and multifunctional design, to obtain a *non-dominated* solution of multiobjective optimization problem (4.1), we perform a *utility-based scalarization* of the multiobjective problem (4.1) as follows:

$$\begin{aligned} \max \quad & \sum_{i=1}^{m} \pi_i U(f_i(\mathbf{x})) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{4.8}$$

where $U : \mathbb{R} \mapsto \mathbb{R}$ is the utility (e.g., "customer satisfaction") corresponding to the design objectives $f_i(\mathbf{x})$, $i = 1, \ldots, m$, that prescribes the preferred order of dominace ("FSD-like", "SSD-like", and so on). Further, $\pi_i \in (0, 1)$ is the relative weight of criterion i, such that $\pi_1 + \ldots + \pi_m = 1$. The weights π_i in (4.8) correspond to probabilities $\mathsf{P}(\omega_i)$ in the expected utility maximization problem (4.7). In a sense, one may consider that the "importance" of design criterion i is derived from the relative "frequency" with which the corresponding property/function of a multifunctional component is utilized during the component's service.

In such a way, the proposed formulation (4.8) represents a novel approach to scalarization in multiobjective optimization, and allows for quantification of multifunctionality. Namely, the solution/design $\mathbf{x}'$ can be considered superior to design $\mathbf{x}''$ with respect to design criteria $f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})$ if

$$\sum_{i=1}^{m} \pi_i U(f_i(\mathbf{x}')) \geq \sum_{i=1}^{m} \pi_i U(f_i(\mathbf{x}'')),$$

and which exploits deep connections to fundamental principles of decision science.

Despite the fact that the proposed multiobjective scalarization method was developed with the goal to be applied for design of multifunctional materials and structures, we were unable to obtain a suitable data set to test our approach in such a setting. Instead, we consider multiobjective extensions of two well-known problems in operations research and decision sciences, namely, a multiobjective shortest path problem, and multiobjective resource allocation problem.

4.3 Computational studies

As an illustration of the proposed framework, two examples are considered: a multi-objective extension of the shortest path problem described in 4.3.1, and resource allocation problem discussed in 4.3.2.

For the purpose of this study the utility function U was chosen to be a logarithmic function, which defines an ordering consistent with stochastic ordering of any degree k (kSD):

$$
\begin{aligned}
&U(x) = \ln x \\
&U'(x) = \frac{1}{x} \geq 0 && \rightarrow \text{FSD} \\
&U''(x) = -\frac{1}{x^2} \leq 0 && \rightarrow \text{SSD} \\
&U'''(x) = \frac{2}{x^3} \geq 0 && \rightarrow \text{3SD} \\
&\quad \dots
\end{aligned}
$$

4.3.1 Multiobjective shortest path problem

For multiobjective shortest path problem (MOSP) one can consider the network where each arc has n attributes with preference given to the smaller values. Such attributes may include, for instance, arc length, the cost of traversing an arc, the "risk" of traversing an arc, etc. Shortest path problem can be easily generalized for the case of multiple objectives. Let s denotes starting node, t denotes sink node, A to be the set of arcs, V is the set of vertices. Let each arc has n cost parameters $c^l_{i,j}$ for $(i,j) \in A$ and $l \in \{1, \dots, n\}$. Then

MOSP can be formulated as follows:

$$\begin{aligned}
\min \quad & (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) \\
\text{s. t.} \quad & \sum_{j:(i,j)\in A} x_{i,j} - \sum_{j:(j,i)\in A} x_{j,i} = \begin{cases} 1 & i = s \\ 0 & \forall i \notin V(\{s,t\}) \\ -1 & i = t \end{cases} \\
& x_{i,j} \geq 0, \quad \forall (i,j) \in A,
\end{aligned}$$

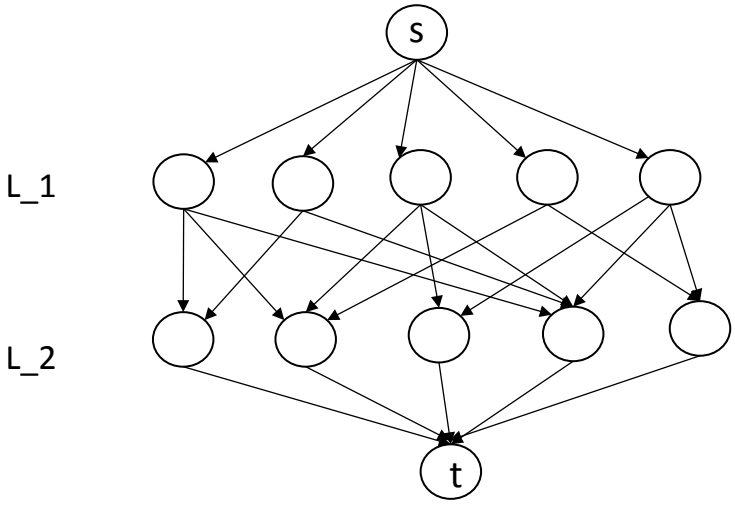
where $f_l(\mathbf{x})$ denotes the path length as measured by l-th attribute:

$$f_l(\mathbf{x}) = \sum_{(i,j)\in A} c_{i,j}^l x_{i,j}, \quad \forall l \in \{1, \ldots, n\}.$$

In our case study, a $(q + 2)$-partite graph was used (with partitions $\{s\}, \{t\}, \{L_1\}, \ldots, \{L_q\}$) constructed in the following way: each of partitions $\{L_i\}$ ($i = 1, \ldots, q$) contains r nodes (total number of nodes in the network is $r^2 + 2r$), starting node s is connected with all nodes in L_1 and only with them, sink node t is connected with all nodes in L_q and only with them, each node in L_i ($i = 1, \ldots, q-1$) is connected only with nodes in L_{i+1}. For simplicity of network construction, density parameter d (denoting density between layers $L_1, \ldots L_q$) was used, so that the number of arcs is $d \times r \times r \times (q-1) + 2r$. Values of parameters $c_{i,j}^l$ were chosen randomly in the range of $[50, \ldots, 100]$; the density parameter d was set to 0.9. In this experiments, the logarithmic function was chosen as the utility function that provides SSD.

Observed results for graphs of different sizes are summarized in Table 1 and Table 2. Table 1 represents values of utility objective $U_F = \pi_i U(f_i(\mathbf{x}))$, weighted sum of objectives

Figure 4.1: Example of the graph used for multiobjective shortest path problem ($r = 5$, $q = 2$)



($S = \sum_{j=1}^{n} f_j/n$) and values of each objective separately (calculated at the optimal point) for utility-based stochastic dominance approach. And for the purpose of comparison Table 2 represents optimization of f_1 solely, here value of U_F was calculated at the optimal point. As one can see, the proposed approach reduces the overall cost while keeping the values of separate objectives reasonably low.

4.3.2 Multiobjective resource allocation problem

Resource allocation problem arises when the decision maker wants to assign available resources in a most reasonable way. One example of such a problem is portfolio optimization.

The problem can be formulates as follows: given a set of assets, one wants to

Table 4.1: Values of functions for utility-based stochastic dominance approach. Here, q is the number of layers L_i in between source and sink nodes, r is the number of nodes in each layer L_i, S is the weighted sum of f_j: $S = \sum_{j=1}^{n} f_j/n$, and U_F is the objective function in proposed method $U_F = \sum_{i=1}^{n} n^{-1} U_i(f_i(x))$

q	r	U_F	S	f_1	f_2	f_3	f_4	f_5	f_6	f_7
5	5	24.93	388	362	372	387	387	416	391	401
5	10	24.66	371.14	385	367	366	377	368	377	358
5	20	25.02	357.71	349	371	375	382	336	352	339
10	5	45.95	725.14	719	726	730	730	715	737	719

Table 4.2: Values of functions for optimization of the attribute f_1 separately.

q	r	U_F	S	f_1	f_2	f_3	f_4	f_5	f_6	f_7
5	5	25.44	425.71	338	423	398	499	453	465	404
5	10	25.39	422.29	321	493	387	441	408	481	425
5	20	25.32	417.86	315	475	398	473	351	452	461
10	5	46.84	794.71	680	828	747	880	816	803	809

make a portfolio with the highest expected return. As expected returns of assets include randomness, one wants to minimize a variance of the resulting portfolio. The set of non-dominated portfolios form an Pareto frontier and there are a number of effective algorithms to construct it. As the main interest of this study is to explore the effectiveness of the proposed method for multiple objectives, it seems reasonable to form the portfolio taking into account the following objectives: expected return, variance, Sharpe ratio, Maximum loss, CVaR. Note that not all of those objectives are for maximization, for example one wants to minimize variance of the portfolio.

Let x_k be a weight of asset k in the portfolio, then the necessary feasibility condition would be $\sum_{k=1}^{n} x_k = 1$. Let ξ_{kj} be a return of asset k under scenario j, $j = 1, ..., N$. Let $R_k = \sum_{j=1}^{N} \xi_{kj}/N$ be an expected return on asset k, and $\sigma_{ij} = \mathsf{covar}(\xi_i, \xi_j)$. With notations just introduced scalar optimization problems for objective functions under consideration would be as follows.

The expected return of the portfolio is maximized:

$$\begin{aligned} \max \quad & \sum_{k=1,...,n} R_i x_i \\ \text{s. t.} \quad & \sum_{k=1,...,n} x_i = 1, \quad x_i \geq 0 \end{aligned}$$

The variance of the portfolio is minimized:

$$\begin{aligned} \min \quad & \sum_{i=1,...,n} \sum_{j=1,...,n} \sigma_{ij} x_i x_j \\ \text{s. t.} \quad & \sum_{k=1,...,n} x_i = 1, \quad x_i \geq 0 \end{aligned}$$

The Maximum loss is minimized:

$$\begin{aligned} \min \quad & M \\ \text{s. t.} \quad & M \geq -\sum_{k=1,...,n} \xi_{kj} x_k, \quad j = 1, \ldots, m, \\ & \sum_{k=1,...,n} x_i = 1, \quad x_i \geq 0 \end{aligned}$$

(here x^j is the "best" possible portfolio selection for the scenario j).

The Conditional Value-at-Risk of portfolio's return is minimized:

$$\begin{aligned} \min \quad & \mathsf{CVaR}_\alpha(-\sum_{k=1,\ldots,n} R_k x_k) = \min_{\eta \in R}\{\eta + \frac{1}{\alpha} E[(-R^T x + \eta)_-]\} \\ \text{s. t.} \quad & \sum_{k=1,\ldots,n} x_i = 1 \\ & x_i \geq 0. \end{aligned}$$

Observe that this problem reduces to an LP.

The Sharpe ratio is maximized:

$$\begin{aligned} \max \quad & \sum_{k=1,\ldots,n} R_k x_k / \sigma \\ \text{s. t.} \quad & \sum_{k=1,\ldots,n} x_i = 1 \\ & x_i \geq 0, \end{aligned}$$

where $\sigma = \sum\sum x_i x_j \sigma_{ij}$.

The corresponding problem of constructing a portfolio that optimizes all the described criteria simultaneously, is formulated as

$$\max \quad F_U = \sum U(f_i(x)) \tag{4.9}$$

$$\text{s. t.} \quad \sum_{k=1,\ldots,n} x_i = 1 \tag{4.10}$$

$$x_i \geq 0 \tag{4.11}$$

where f_i are expected values of portfolio returns for objective functions described above, and all criteria have equal weights ($\pi_i = 1/m$).

The case studies were performed using data of S&P500 companies, namely daily closing prices from January 1, 2005 to September 30,2015. Only cases with full knowledge

of historical prices were left in the data pool, i.e. the total number of assets to choose from was 483, and the total number of possible scenarios 2760. For the case study assets and scenarios were chosen randomly.

Table 4.3: Expected returns for utility-based stochastic dominance approach for Resource allocation problem. U_F is the objective function in proposed method $U_F = \sum_{i=1}^{n} U_i(f_i(x))$

n	N	F_U	SharpeRatio	MaxLoss	ExpectedReturn	MinVariance	CVaR
10	10	8.31	62.68	62.68	62.68	8.31	64.34
10	20	7.39	63.56	63.56	63.56	7.39	65.04
10	30	10.59	66.07	66.07	66.07	10.59	68.18
20	10	0.28	274.69	274.69	274.69	0.28	274.74
20	20	0.21	264.30	264.30	264.30	0.21	264.34
20	30	0.14	255.26	255.26	255.26	0.14	255.29
30	10	0.25	274.69	310.47	310.47	0.25	281.89
30	20	0.19	264.30	309.32	309.32	0.19	273.34
30	30	0.15	255.26	302.70	302.70	0.15	264.78
40	10	0.31	274.69	310.47	310.47	0.31	281.90
40	20	0.35	264.30	309.32	309.32	0.35	273.37
40	30	0.22	255.26	302.70	302.70	0.22	264.79
50	10	4.13	274.69	310.47	310.47	1.21	282.67
50	20	0.52	264.30	309.32	309.32	0.52	273.40
50	30	0.29	255.26	302.70	302.70	0.29	264.81
60	10	5.58	274.69	310.47	310.47	0.41	282.96
60	20	0.50	264.30	309.32	309.32	0.50	273.40
60	30	0.27	255.26	302.70	302.70	0.27	264.80

4.4 Conclusions

This study explores the new method for multiobjective optimization, which is based on the stochastic dominance concept. In case of multiple objectives it is not always possible to choose a solution that optimizes all the objectives, for example, it is not always possible

to minimize the cost and maximize efficiency of the system at the same time. Instead, one has to find an acceptable tradeoff between those two objectives. Multiple techniques are known for finding such tradeoff, one of them is Pareto dominance, when a solution is called non-dominated if it is not possible to improve one of the objectives without worsen at least one of the others. Similar to Pareto optimality the concept of Stochastic Dominance is used in this work to make a meaningful comparison between objective values. Also, the well-known connection between stochastic dominance and the utility theory was used for scalarization of the vectorial objective.

CHAPTER 5
OPTIMIZATION TECHNIQUES FOR CRYSTAL STRUCTURE PREDICTION BASED ON X-RAY CRYSTALLOGRAPHY DATA

5.1 Introduction

Structure of the organic molecule is essential for understanding properties and potential advantages of a given material. As it is not always possible to derive the structure of organic crystal from its chemical composition, analysis of X-ray diffraction data becomes an essential part of crystallographic science. Such an analysis is usually conducted as a two-step procedure, where the structure is suggested in the first step, and an iterative procedure of improving the initial structure is performed afterwards; the latter procedure is called refinement. In the refinement phase the minimum energy principle is used along with structural constraints to obtain the degree of correlation between the proposed structure and the observed data, with the goal of minimizing the mismatches. Although finding correlation at the refinement step typically involves complicated computation, it is a well-known area with a number of approaches developed in the literature. Nevertheless, the initial guess about the structure and subsequent improvement require experienced operator, and are usually done by trial-and-error method. The project described in this chapter is focused on automatization of the whole crystal structure analysis procedure, so the problem can be formulated and solved without human involvement. The corresponding optimization problem is formulated as a minimization of the refinement value over all possible crystal structures. Since estimation of refinement value is well-developed area, the problem can be treated as black-box combinatorial optimization problem, where the value of the objective

function for a given selection of decision variables is provided by a “black box” device. For large crystals, a straightforward enumeration becomes very resource- and time demanding, so heuristic optimization techniques were used.

5.2 X-ray optics in crystallography

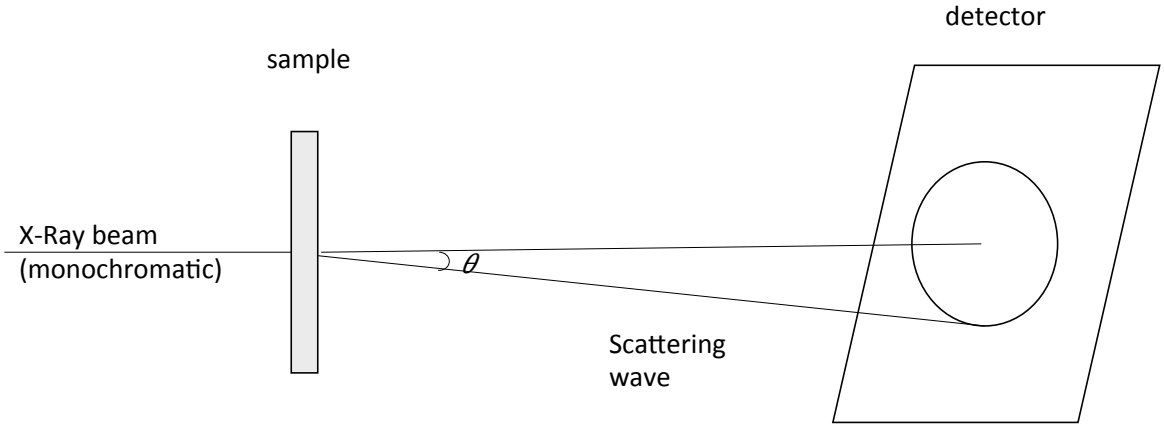
In 1914, a German physicist Max von Laue won a Nobel prize in Physics “For his discovery of the diffraction of X-rays by crystals”. His research was critical for development of X-ray spectroscopy. It was shown that crystals scatter X-ray radiation in different patters, which carry the information about crystal composition (structure). However, X-ray crystallography is not suitable for non-crystalline solids, and since many substances do not form solid crystals (for example, proteins), the Small-Angle X-ray Scattering (SAXS) was developed. SAXS is a technology that allows for detection of inhomogeneous electron density at scales of 10 to 1000 Angstroms [17].

The SAXS experiment can be described as follows: the sample is placed between an X-ray source and a detector, and a focused X-ray beam passes through the sample. The nonhomogeneous electron density in the sample crystal causes the scattering pattern, which can be observed in the form of intensity distribution.

Most approaches to X-ray scattering are based on the basic principals of quantum mechanics and electromagnetic interactions, where electrons are considered as classical oscillators [7, 12, 42].

Electron is said to scatter X-rays when under X-ray radiation the electron is set into continuous oscillation and emits an electromagnetic wave. It is the case of elastic scattering,

Figure 5.1: A schematic representation of SAXS experiment



as the scattered beam and the incident beam are coherent (have the same frequency). The dependence of the intensity of scattered beam on the angle of scattering is described by Thomson's equation [12]:

$$I = I_0 \frac{e^4}{r^2 m^2 c^4} \frac{1 + \cos^2 2\theta}{2},$$

where I is the intensity of the scattered beam, I_0 is the intensity of the incident beam, r is the distance between the electron and the detector, θ is the angle of scattering, and e, m, and c are the standard constants: e is the electron charge, m is the mass of the electron, and c is the speed of light in vacuum.

When X-ray is scattered by an atom, it is modeled as a superposition of scattering by each electron of the atom, while nuclei scattering can be considered as negligible as

the mass of the nuclei is much larger relative to the mass of the electron. However, when X-ray beam heats the atom, electrons move in different positions, which causes the loss of coherence property and therefore the intensities cannot be summed directly. Because of the phase difference for the scattered waves, a partial interference occurs and that decreases the net intensity. To describe the efficiency of scattering by an atom in a given direction, the following ratio can be introduced:

$$f = \frac{\text{amplitude of the wave scattered by an atom}}{\text{amplitude of the wave scattered by one electron}},$$

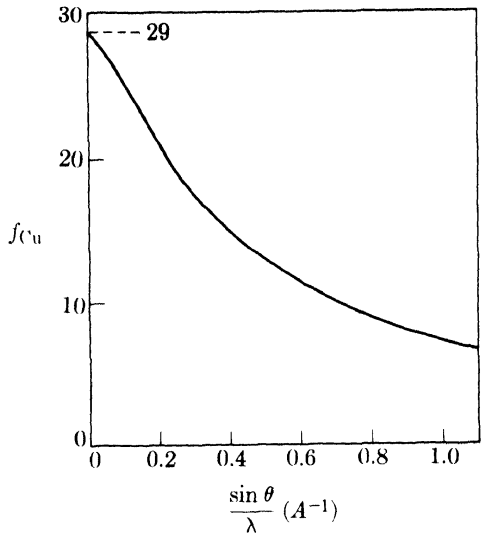
which is called the atomic scattering factor. For $\theta = 0$, when all scattered waves are coherent, the atomic scattering factor would be equal to the number of electrons of the given atom. For $\theta \neq 0$, the value of f decreases with increase in the value of θ. Moreover, θ is reciprocal to $\sin\dfrac{\theta}{\lambda}$, where λ is the wave length of an incident beam. This effect is illustrated in Figure 5.2. As the intensity of the wave is proportional to amplitude squared, the intensity I of the scattered beam satisfies $I \sim f^2$.

Substance can be considered as a periodic arrangement (one unit is called unit cell) of atoms, and its scattering pattern depends on the chemical composition, the distances between atoms, type of crystal symmetry, and so on (which are collectively known as the internal structure). The result of X-ray scattering by unit cell is described by the structure factor F:

$$F_{hkl} = \sum_{n=1}^{N} f_n e^{2\pi(hu_n+kv_n+lw_n)i},$$

where N is the total number of atoms in the unit cell, f_n is the scattering factor of each atom, (u_n, v_n, w_n) are internal coordinates of the atom n, and (h, k, l) are axes unit vectors.

Figure 5.2: Cooper atomic scattering factor dependence on scattering angle [12]



F describes both the phase and the amplitude of the scattered wave, so the intensity of the beam would be proportional to F^2, namely

$$I = F^2 p \frac{1 + \cos^2 2\theta}{\sin^2 \theta \cos \theta},$$

where p is multiplicity factor, i.e. the number of different planes in a form having the same spacing.

5.2.1 Qualitative analysis of the crystal structure

One of main interests of X-ray crystallography is to perform qualitative analysis, i.e., to determine the pattern of the substance under consideration. As it was stated before, the intensity distribution is proportional to the structure factor squared, but the phase information is lost, which makes it impossible to derive the electron density map directly from the X-ray diffraction information. To overcome this difficulty, the two-stage solution

process is used.

At the first stage, one makes an assumption about the possible structure. The initial assumption can be based on the internal knowledge of the crystal composition, or the expected structure of the molecule (if it was engineered), or any other information one might possess about this particular crystal. At the second stage, called the refinement process, an iterative procedure of improving the assumed structure is conducted. At each iteration the assumed structure is updated with hope to match the experimental data and the calculated pattern of intensities of the suggested structure (see, for example, [44]). Therefore, at each iteration one would need a quantitative representation of the correlation between the assumed and the experimentally observed data, which is represented by the *fitness function*. Such a function is typically expressed as correlation between structure factors [24, 36]:

$$R = \frac{\sum_{\text{all reflections}} (F_0 - F_c)}{\sum_{\text{all reflections}} (F_c)},$$

where F_c is the calculated structure factor and F_0 is the structure factor mathematically obtained from the observed intensities I. The lower value R is, the better the suggested structure fits the experimental data. Sometimes, a weighted fitting function is used:

$$R = \left(\frac{\sum_{\text{all reflections}} w_i (F_0 - F_c)^2}{\sum_{\text{all reflections}} (F_c)^2} \right)^{1/2}, \tag{5.1}$$

where $w_i \geq 0$ are the weight factors, $\sum w_i = 1$.

5.3 Solution methods

5.3.1 Problem formulation and variable description

Suppose that one is trying to determine the structure of an organic crystal with known chemical composition, i.e., the types of elements and the number of them in the

crystal, and symmetry class. As an output of the crystallographic experiment, one can obtain a file containing the following components: $(h, k, l, I, \sigma^2(I))$ where (h, k, l) are spacial coordinates of the intensity peaks, I is the relative value of intensity, and scattering angle. Given this data, the problem is to assign each intensity peak to a certain element, so that the fitting function (5.1) is minimized.

Peaks with higher intensities are more likely to be elements (rather then noise), so in crystallographic data all peaks are sorted in descending order, with highest intensity peaks coming first. Also, as the number of output intensity peaks in crystallographic file is large and includes every minimal intensity fluctuation that was detected, the total number of peaks considered is not larger then the total number of elements of non-hydrogen elements plus half of the number of hydrogen elements. This constraint is valid as hydrogen is hard to detect due to its low periodic number.

Let the string $(Q_1, ..., Q_n)$ represent certain chemical composition, variables Q_i can take values from the chemical composition element set. For example, if chemical composition is "C_2H_5OH", variables Q_i can take values C, O, H and $NONE$ ($NONE$ means that this peak does not correspond to any element, i.e., it represents noise), the total number of variables $n = 2 + 1 + (5 + 1)/2 = 6$, the total number of different elements is $m = 3 + 1 = 4$. The fitting function $F = F(Q_1, ..., Q_n)$ is calculated using external software (black box).

5.3.2 Combinatorial Problem Formulation

Let the crystal (chemical composition) contain m different elements, with the amount of each element being denoted by d_i, $i = 0, \ldots, m$ (we let the 0-element to denote hydrogen). From the crystallographic information one can obtain first n peaks of highest intensity (where $n = d_1 + \ldots + d_m$ is the total number of non-hydrogen atoms in the crystal). The problem is to find an assignment $\mathbf{x}$ of intensity peaks to elements that minimizes the correlation function $F(\mathbf{x})$. In this case, the objective function $F(\mathbf{x})$ is an unknown function, $F(\mathbf{x}) \in [0, 1]$. The the binary variables $\mathbf{x}_{ij} = 1$ denote assignment of peaks to elements, i.e.,

$$x_{ij} = \begin{cases} 1, & \text{if peak } i \text{ is assigned to element } j\ , \quad i = 1, \ldots, n,\ j = 1, \ldots, m, \\ 0, & \text{otherwise.} \end{cases}$$

Then, the problem can be formulated as the following assignment problem:

$$\min \quad F(\mathbf{x}) \tag{5.2a}$$

$$\text{s. t.} \quad \sum_{i=1}^{m} x_{ij} = 1 \quad \text{for} \quad j = 1, \ldots, n \tag{5.2b}$$

$$\sum_{j=1}^{n} x_{ij} \leq d_i \quad \text{for} \quad i = 1, \ldots, m \tag{5.2c}$$

$$\sum_{i=j}^{n} x_{ij} = d_i. \tag{5.2d}$$

Constraints (5.2b) ensure that each peak is assigned to only one element, and constraints (5.2c) ensure that the capacity of elements will not exceeded the prescribed values, while constraint (5.2d) ensures that each element is assigned to some peak.

5.3.3 Nearest neighbor search

For the problem formulation given above, the simple Nearest Neighbor (NN) search can be performed. Let solution $(A_1, ..., A_n)$ be called 1-neighbor of solution $(B_1, ..., B_n)$ if there exists such integer number k that $A_i = B_i$ for all $i = \{1, ..., n\}\backslash\{k\}$ and $A_k \neq B_k$. In other words, 1-neighbor solutions differ in just one position. Let set $\{E_1, ..., E_m\}$ be the numerated set of chemical elements to use with $E_1 = C$, $E_{m-1} = H$ and $E_m =$ "$NONE$"; let $V = (v_1, ..., v_m)$ denote the maximum number of elements in the molecule, i.e. number of element E_1 is v_1 and so on.

Then, at each iteration the Nearest Neighbor Search algorithm (see Algorithm 5.1) inspects all possible 1-neighbor solutions in order to improve current solution. If such an improvement is possible, the best improved solution becomes the current solution and the algorithm proceeds to the next iteration. Algorithms terminates when the maximum number of iterations $N_{maximum}$ is achieved, or when no improvement is made during the current iteration.

If the initial solution for Algorithm 5.1 is chosen as $x_0 = (Q_1, ..., Q_n) = (C, ..., C)$, and carbon C is excluded from the set $\{E_1, ..., E_m\}$, at each iteration the solution would be constantly improved and no two solutions from different iterations would coincide (as number of carbon elements on each iteration is different). Also, one can note that if the initial solution x_0 in Algorithm 5.1 is chosen to be feasible, at each iteration feasibility preserves, so the x_{best} would be feasible as well.

There are several possibilities for outcome of Algorithm 5.1:

- Ideally, when Algorithm 5.1 terminates, one would have the vector $U =$

Algorithm 5.1 General Nearest Neighbor search algorithm

Set the initial solution to be $x_0 = (Q_1, ..., Q_n) = (C, ..., C)$, calculate $F_0 = F(x_0)$

Set the vector $U = (v_1 - n, ..., v_m)$ to be the current number of elements available to use

Improvement := true

Iteration $= 0$

while Improvement $=$ true && Iteration $<= N_{maximum}$ **do**

Iteration$+ = 1$

Improvement := false

$x_{iteration} = x_{best}$

for $i = 1$ **to** n **do**

for $j = 2$ **to** m **do**

if $u_j > 0$ **then**

Calculate $F_1 = F(Q_1, ..., Q_{i-1}, E_j, Q_{i+1}, ..., Q_n)$

if $F_1 <= F_{best}$ **then**

$x_{best} = (Q_1, ..., Q_{i-1}, E_j, Q_{i+1}, ..., Q_n$

$F_{best} = F_1$

Improvement := true

end if

end if

end for

if Improvement = true **then**

$u_1 + = 1$

end if

end for

end while

$(0, 0, ..., 0, u_{n-1}, u_n)$ as variables u_{n-1} and u_n correspond to the hydrogen H and $NONE$ elements. If this is not the case, the adjustment algorithm 5.2 should be performed.

- Vector U returns correct numbers, but the solution is not optimal. This case is possible when two elements with close periodic numbers (such as, for instance, C and N) switched places, i.e., 2-neighbor of the solution would be optimal. To avoid such situations, the check Algorithm 5.5 should be performed.

Algorithm 5.2 Nearest Neighbor adjustment algorithm

Initial solution is $x = (Q_1, ..., Q_n)$, $F = F(x_0)$, $U = (u_1, ..., u_m)$

for $j = 1$ **to** m **do**

if $(u_j < 0)$ **then**

Run Algorithm 5.4

end if

if $u_j > 0$ && $(m \leq m - 2)$ **then**

Run Algorithm 5.3

end if

$x = x_1$, $F = F_1$

end for

Algorithm 5.4 finds feasible points for those elements that are "overused," and Algorithm 5.3 finds feasible points for "underused" elements. Note that the result of Algo-

Algorithm 5.3 Nearest Neighbor low adjustment algorithm

$F_1 = 1$

for $i = 1$ **to** n **do**

if $Q_i =$ "$NONE$" **then**

Calculate $F_2 = F(Q_1, ..., Q_{i-1}, E_j, Q_{i+1}, ..., Q_n)$

end if

if $F_2 <= F_1$ **then**

$x_1 = (Q_1, ..., Q_{i-1}, \text{“}NONE\text{”}, Q_{i+1}, ..., Q_n), \quad F_1 = F_2$

end if

end for

Algorithm 5.4 Nearest Neighbor high adjustment algorithm

$F_1 = 1$

for $i = 1$ **to** n **do**

if $Q_i = E_j$ **then**

Calculate $F_2 = F(Q_1, ..., Q_{i-1}, \text{“}NONE\text{”}, Q_{i+1}, ..., Q_n)$

end if

if $F_2 <= F_1$ **then**

$x_1 = (Q_1, ..., Q_{i-1}, \text{“}NONE\text{”}, Q_{i+1}, ..., Q_n), \quad F_1 = F_2$

end if

end for

Algorithm 5.5 2-Neighbor check algorithm

Initial solution is $x_0 = (Q_1, ..., Q_n)$, $F_0 = F(x_0)$

Set $F_{best} = F_0$, $x_{best} = x_0$

Iteration $= 0$

while Iteration $<= N_{maximum}$ **do**

Iteration$+ = 1$

Improvement := false

$x_{iteration} = x_0$

for $i = 1$ **to** n **do**

for $j = 1$ **to** n **do**

Calculate $F_1 = F(Q_1, ..., Q_{i-1}, Q_j, Q_{i+1}, ..., Q_{j-1}, Q_i, Q_{j+1}, ..., Q_n)$

if $F_1 <= F_{best}$ **then**

$x_{best} = (Q_1, ..., Q_{i-1}, Q_j, Q_{i+1}, ..., Q_{j-1}, Q_i, Q_{j+1}, ..., Q_n)$

$F_{best} = F_1$

end if

end for

end for

end while

rithm 5.2 might be far from optimal and one might need a new run of Algorithm 5.1 with adjusted initial solution.

As Algorithms 5.3 and 5.4 are introduced, one might construct a variation of the Nearest Neighbor Search (Algorithm 5.6) that allows for infeasible solutions during p iterations. Note that Algorithm 5.6 allows one to check the same solution at different iterations, therefore its computational cost would be higher than that of Algorithm 5.1, but the obtained solution is expected to be better.

5.3.4 Simulated annealing

Simulated Annealing (SA) is a global optimization heuristic that is inspired by annealing process in metallurgy. Originally it was introduced by Scott Kirkpatrick in 1983 [27] for combinatorial optimization problems, namely the Traveling Salesman Problem. One of the advantages of the SA is that it does not get "stuck" at local minima, as by its nature the SA algorithm accepts solutions with worse objective function with nonzero probability. The general SA algorithm is outlined below [4].

Let $f : S \to R$ be an objective function to be minimized that is defined on a finite set S. For each element $s \in S$ there is a set $N(s) \in S$ that is called neighborhood of s, for any two elements $s_1 \in S$ and $s_2 \in S$, $s_1 \in N(s_2)$ if and only if $s_2 \in N(s_1)$. Let $x_0 \in S$ to be initial solution, $f(x_0)$ is initial value of the objective function. As stopping criteria of the algorithm one can choose the maximum number of steps, or threshold on change in objective value δ: $|f(x_t) - f(x_{t+1})| \leq \delta$, or any other criterion that is suitable for the problem. The general SA algorithm is formalized in Algorithm 5.7.

Algorithm 5.6 Infeasible variation of Nearest Neighbor search algorithm

$x_0 = (Q_1, ..., Q_n) = (C, ..., C)$,$F_0 = F(x_0)$, $U = (v_1 - n, ..., v_m)$

Improvement := true

Iteration $= 0$

while Improvement $=$ true **and** Iteration $<= N_{maximum}$ **do**

Iteration $=$ Iteration $+ 1$

Improvement := false

for $i = 1$ **to** n **do**

for $j = 1$ **to** m **do**

Calculate $F_1 = F(Q_1, ..., Q_{i-1}, E_j, Q_{i+1}, ..., Q_n)$

if $F_1 <= F_{best}$ **then**

$x_{best} = (Q_1, ..., Q_{i-1}, E_j, Q_{i+1}, ..., Q_n)$, $F_{best} = F_1$

Improvement := true

end if

if REMINDER(Iteration$/p$)$= 0$ **then**

Run Algorithm 5.2

end if

end for

if Improvement = true **then**

$u_1 += 1$

end if

end for

end while

Run Algorithm 5.2, Algorithm5.5

Algorithm 5.7 General SA algorithm

Set t:=0, choose x_0 and calculate $f(x_0)$;

while Stopping criteria are not satisfied **do**

Choose $y \in N(x_t)$ randomly;

Set $x_{t+1} := y$ with probability $\min(1, e^{f(x_t)-f(y)/T(t)})$ and $x_{t+1} := x_t$ otherwise

Set $t := t + 1$

end while

The function $T(t) : N \to R^+$ is called *cooling schedule* and is usually set as

$$T(t) = \frac{d}{\log t},$$

where d is some positive constant, $d > 0$.

Simulated Annealing has been extensively studied in the literature, and has been found to possess many important properties, including the fact that it is guaranteed to converge to an optimal solution, which is especially crucial in the context of non-convex problems:

Theorem 4 ([18]). *We say that state s communicates with optimal set S^* at height h if there exists a path in S (with each element of the path being a neighbor of the preceding element) that starts at s and ends at some element of S^* and such that the larges value of f along the path is $f(s) + h$. Let d^* be the smallest number such that every $s \in S$ communicate with S^* at height d^*. Then, the SA algorithm converges if and only if* $\lim_{t\to\infty} T(t) = 0$ *and*

$$\sum_{t=1}^{\infty} e^{-d^*/T(t)} = \infty.$$

SA algorithm is used in crystallography in many ways: to predict possible crystal structures given chemical composition, see, for example [14, 35]; for calculation of fitting function in refinement stage [11]; for structural determination when structural fragments are a priory known [45].

5.3.5 Genetic algorithm

Genetic Algorithm (GA) is another nature-inspired optimization metaheuristic, which had found many applications in different areas of science and engineering [23, 25]. It is based on Darwin's principle of "survival of the fittest" and implements three basic ideas:

- If the individual is above-average, its genes will pass to the offspring with higher probability
- If the individual is below-average, its genes will pass to the offspring with lower probability, and will have less affect on the population
- Occasional mutations might happen, and result of mutation might be advantageous or disadvantageous.

GA is an iterative algorithm, and at each iteration it maintains a set of solutions, known as *population.*

Let $F : S \to R$ be the objective function, also called fitting function. Let $x_i \in S$ be an individual solution, in this case it would be a string of a certain length. A *crossover* operation is defined for any pair of solutions as an exchange of portions of those solutions, for example, in general case, if $a = (a_1, a_2, ..., a_n)$ and $b = (b_1, b_2, ..., b_n)$ are

two chosen individuals, then the result (offspring) of most common crossover would be $(a_1, a_2, ..., a_k, b_{k+1}, ..., b_n)$ and $(b_1, b_2, ..., b_k, a_{k+1}, ..., a_n)$, where $k \in [1, n]$ is an integer number chosen at random. A mutation operation is defined for any individual solution as random change of portion of that solution.

The initial population is often chosen at random, and at each iteration of GA the population is updated as follows:

1. Each individual in the population is evaluated with fitness function, and the fittest individual is recorded
2. Portion p_s of population is carried to the next iteration without any change
3. Portion p_c of population is used with crossover operation and its offspring is carried to the next iteration
4. Portion p_m of population is used with mutation operation and the result is carried for the next iteration.

The parameters p_s, p_c, and p_m above satisfy $p_s + p_c + p_m = 1$. Parameters p_s, p_c, p_m may be constant, or may depend on the iteration, for example p_m is often selected to decrease as algorithm proceeds for better convergence. GA iterates until termination criteria are satisfied. As termination criteria one can use the maximum allowable number of iterations, or threshold on fitness function, or any other suitable condition.

In crystallography, Genetic Algorithms are used for defining better unit cell indexing [26], solving phase problem [1], and determining structure with the use of structural components [19]

5.4 Experimental results

The computational studies were conducted using data for several new organometallic crystals of various sizes. The obtained results were compared to the previously known solutions (configurations), obtained using human expertise. We refer to these previously known solutions as the *benchmark solutions*. There is a number of factors affecting the value of objective function, such as the assignment of hydrogen atoms (but in most cases it might be omitted), finding more precise atom positioning, the use of different lattice symmetry class, and so on. We call a solution *acceptable* if its atom configuration coincides with the benchmark solution while the fitting function values might be slightly different. For all algorithms, running time limit of 20 minutes was set. For black-box objective function evaluation, SHELX software was used. Results are summarized in Tables 5.1 and 5.2.

For the problem stated in Section 5.3.1, three algorithms were implemented: Nearest Neighbor search (Algorithms 5.1 and 5.6), Simulated Annealing (Algorithm 5.7), and Genetic Algorithm.

For Nearest Neighbor search it was found that Algorithm 5.6 is less time-efficient, but it returns the feasible optimal solution. Algorithm 5.1 with adjustment made to the final solution, but without check algorithm is faster, but the quality of solution is slightly lower. For the comparison of algorithms Algorithm 5.6 was used.

For SA algorithm definition of 1-neighborhood from Section 5.3.3 was used along with the cooling schedule

$$T_t = \frac{c}{1 + t\sqrt{c}},$$

where constant c denotes the percentage of non-carbon, non-hydrogen atoms in the compound.

For GA, the population of 100 solutions was used. Due to the large number of objective function evaluations, this algorithm was found to be less efficient. GA converges to an optimal solution for each instance, but only in a few of instances an optimal solution was constructed within the predefined time limit.

Schematic examples of structures are shown in Figures 5.3–5.8. In these examples, two structure were chosen: N-2,4-dibromophenylpyridinium chloride (denoted as "fcp1415" in computational results) and 4,4′ bis (N-3-iodopyridinium) tetraphenylethylene bromide (denoted as "fcp157"). Figures 5.3 and 5.4 show the initial and final peak assignments for fcp1415, note that solid lines in the mentioned figures denote covalent bounds. Figures 5.5 and 5.6 show the initial and final peak assignments for fcp157.

The computational results of objective function look promising for organic crystal structures. The next step in this research would be to extend the proposed algorithms and test them on non-organic crystal structures. In addition, it would be of interest to incorporate heuristics based on chemistry laws and relations into the standard metaheuristic methods, such as NNNand implement more of chemistry knowledge in the algorithms.

5.5 Conclusions

This study explored the use of optimization techniques for qualitative analysis of organic crystals. Given X-ray diffraction data of the compound, one can predict the structure of the crystal that correlates best with intensity distribution picture. Such prediction

Table 5.1: Comparison of fitting function values for solutions obtained using Simulated Annealing algorithm (SA), Nearest Neighbor search (NN), and the benchmark solution

Crystal	SA	NN	Benchmark solution	Sample Contains
fcp138	0.1	0.12	0.1	C196 H144 N8 O54
fcp139	0.03	0.05	0.0322	C52 H36 N4 O4
fcp143	0.045	0.07	0.0444	C304 H240 N16 Br32
fcp144	0.055	0.054	0.0584	C152 H120 N8 Br16
fcp148	0.07	0.9	0.0733	C44H36 N4 Cl4 I4 O4
fcp149	0.13	0.18	0.1565	C80 H72 N8 Cl8 Br8
fcp1311	0.13	0.16	0.1893	C92 H100 N8 O48 Zn2
fcp1314	0.068	0.07	0.0716	C56 H44 N4 O4
fcp1316	0.12	0.14	0.1348	C12 H11 N1 Br1 I1
fcp1410	0.027	0.03	0.0247	C40 H32 N4 Cl4 Br8
fcp1411	0.03	0.035	0.0294	C88 H72 N8 O8 Cl8 Br8
fcp1412	0.03	0.03	0.0259	C88 H72 N8 O8 Cl8 I8 O8
fcp1415	0.032	0.033	0.0323	C44 H32 N4 O4 Cl4 Br8
fcp157	0.049	0.052	0.05	C40 H36 N2 I2 Br2

is usually done by an expert with deep knowledge of crystal composition. In previous studies, few attempts to use optimization techniques to find the fittest structure were done, instead they were focused on structure reconstruction from a previously known fragments [45],[19]. In this work, it was assumed that the only knowledge about the structure is its chemical composition and X-ray data.

In general, one can enumerate all possible variations of intensity peaks assignment to elements of chemical composition to get the best structure, but due to the large dimensionality of data such a “brute-force” method would not be efficient. This work discussed the usage of few heuristic techniques for better crystal structure prediction. Experimental results show that all three algorithms (NN, SA and GA) converged to the desired solution. It

Table 5.2: Comparison of running time (in seconds) for Simulated Annealing algorithm (SA), Nearest Neighbor search (NN) and Genetic Algorithm (GA)

instance	SA	NN	GA
fcp138	1108	1205	-
fcp139	332	358	1169
fcp143	1071	1051	-
fcp144	984	902	-
fcp148	582	361	1094
fcp149	973	1007	-
fcp1311	820	1015	-
fcp1314	347	462	1054
fcp1316	287	264	731
fcp1410	659	731	-
fcp1411	762	940	-
fcp1412	896	913	-
fcp1415	637	704	-
fcp157	604	596	1083

was found that GA failed to reach termination criterion within the allowed computational time, whereas NN and SA converge to a solution reasonably fast. The crystal structures used for this study are organic crystals with metallic elements included, and one of the possible extensions of this project would be to generalize the proposed algorithms for other types of crystals from organic and inorganic chemistry.

Figure 5.3: N-2,4-dibromophenylpyridinium chloride (fcp1415) initial structure

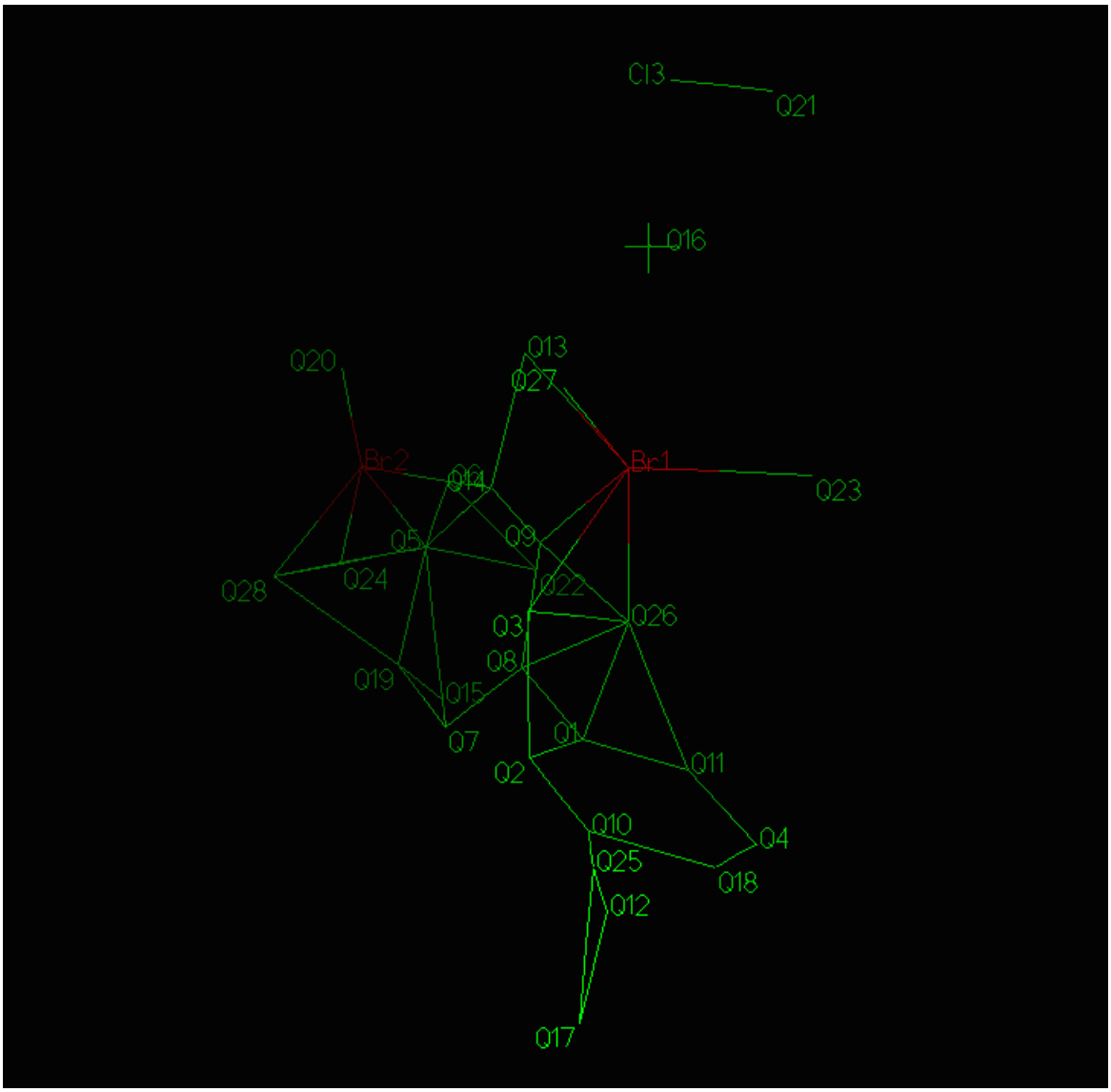


Figure 5.4: N-2,4-dibromophenylpyridinium chloride (fcp1415) solved structure

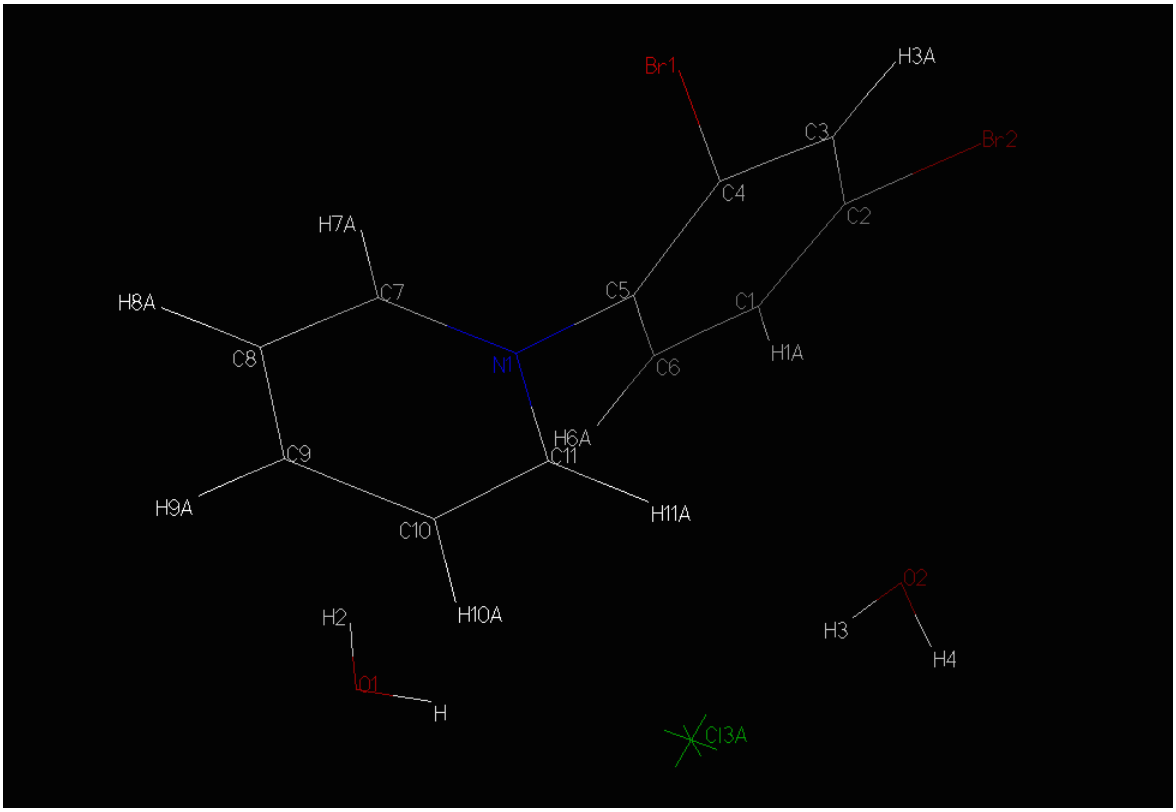


Figure 5.5: 4,4’ bis(N-3-iodopyridinium) tetraphenylethylene bromide (fcp157) initial structure

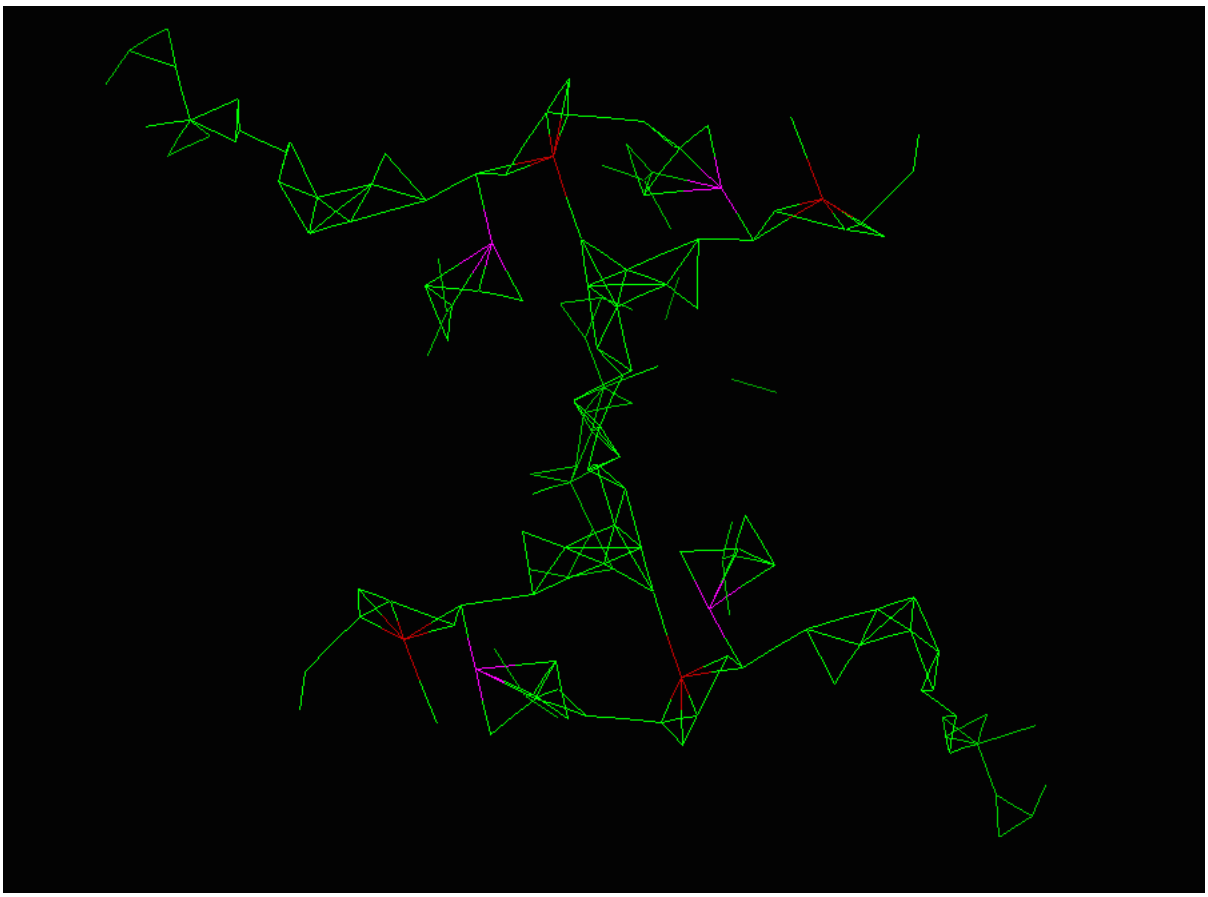


Figure 5.6: Whole solved structure of 4,4’ bis(N-3-iodopyridinium) tetraphenylethylene bromide (fcp157)

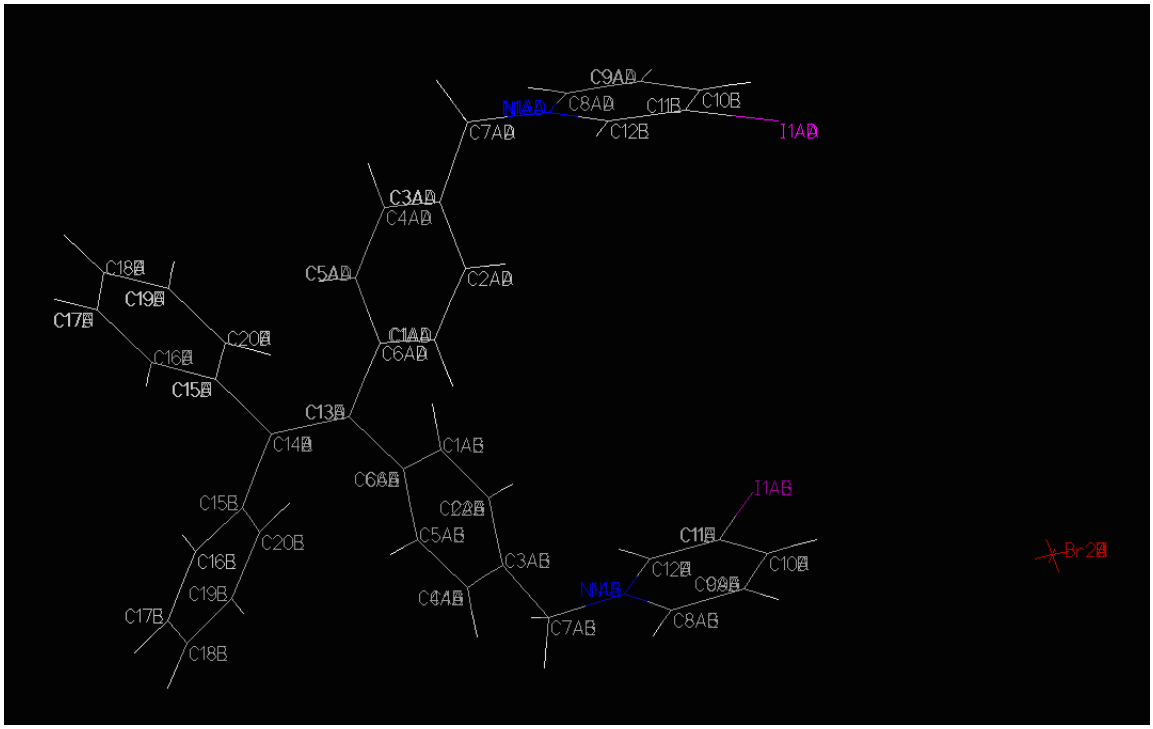


Figure 5.7: Example of asymmetric unit for 4,4’ bis(N-3-iodopyridinium) tetraphenylethylene bromide (fcp157)

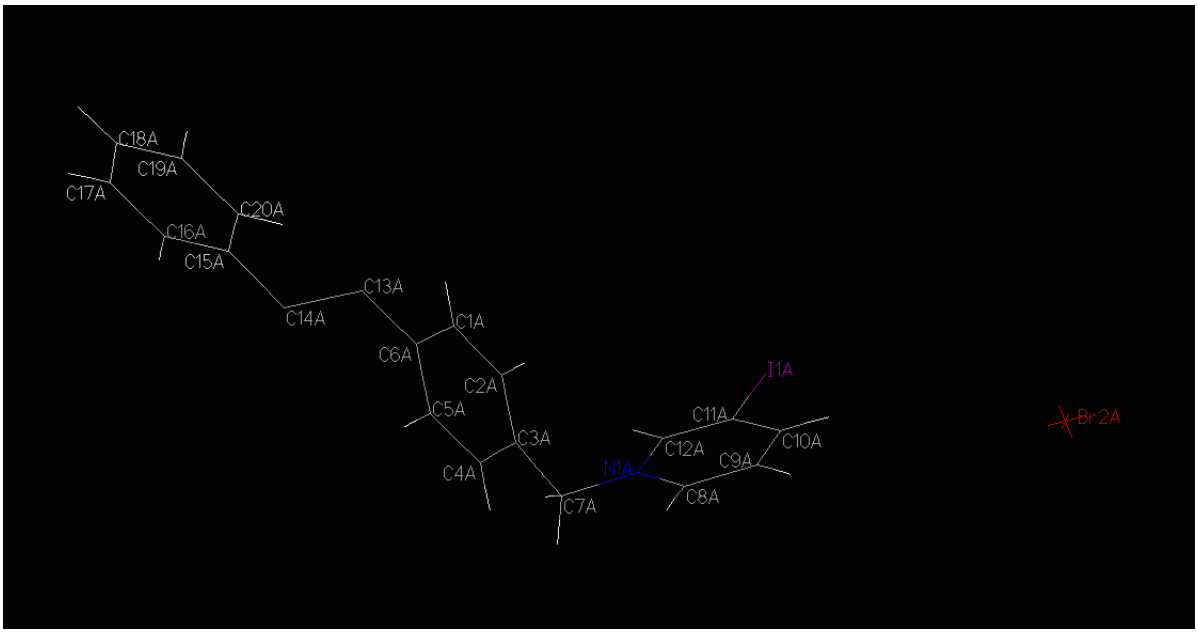
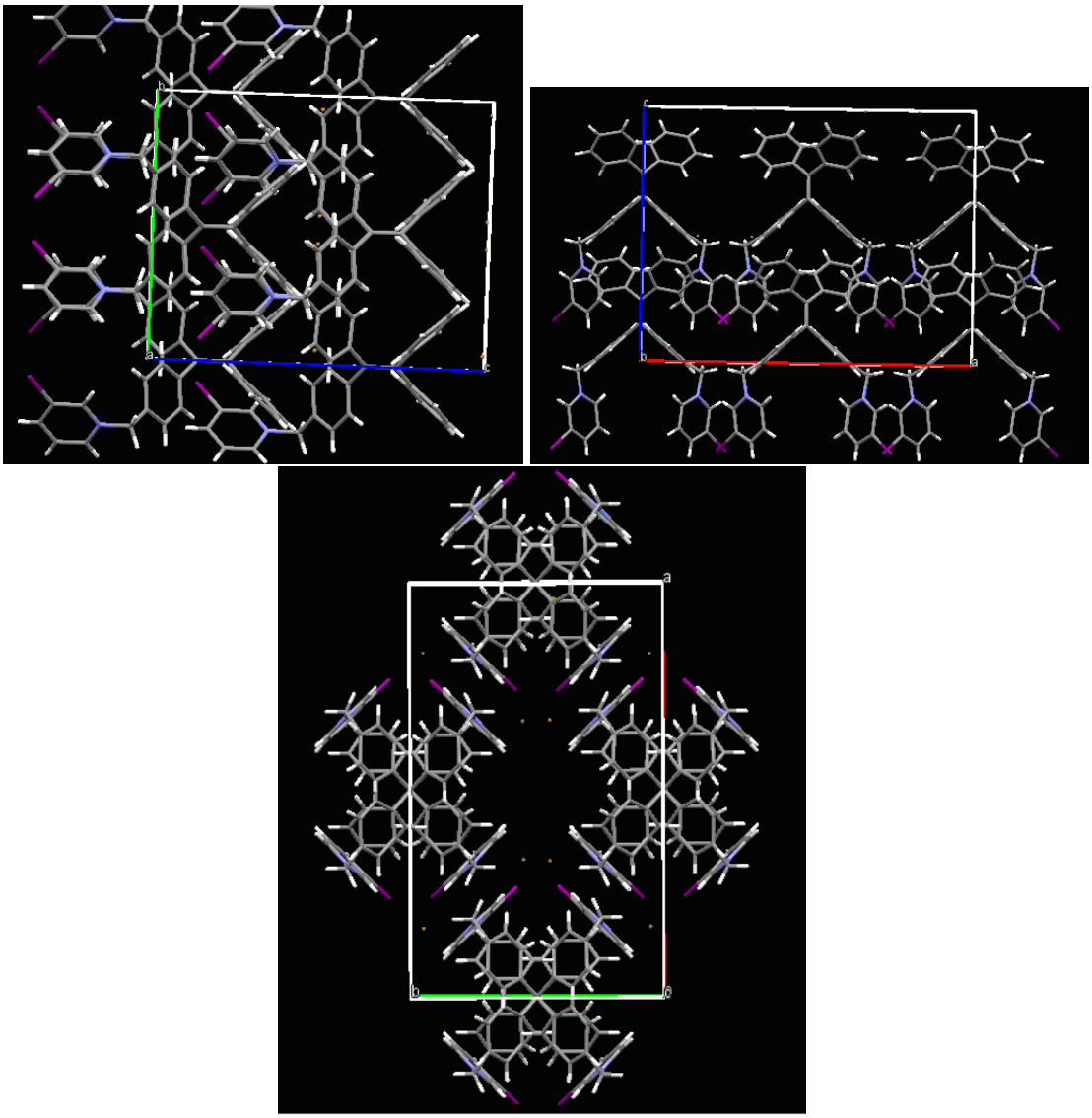


Figure 5.8: View down different axises for 4,4' bis(N-3-iodopyridinium) tetraphenylethylene bromide (fcp157)



CHAPTER 6
CONCLUSIONS

This work consists of four projects, each related to the problems arising in Industrial Engineering and Material Design. Although all chapters seem to be relatively disconnected, all of them can be applied in the Material Design area of study.

Chapter 2 is focused on the new model of linear classification, base on p-order conic programming. This technique was tested on real-life datasets related to biomedical data. The possible expansion of this work would be the application of proposed method for classification of materials.

In Chapter 4 the new technique for multiobjective optimization is presented. Case studies for this technique are multiobjective shortest path problems and portfolio optimization. This study can be applied for design of multifunctional materials.

Chapter 3 introduces an optimization approach to determine the tightest possible bounds for effective overall elastic moduli of composite materials. In future, the solution for general distribution of inclusions case is to be found.

Chapter 5 discussed the qualitative analysis of X-ray data using heuristic techniques. The expantion of this work would be to explore implementation of proposed approach for inorganic structures.

REFERENCES

[1] I. Abdurahman and A. Purwanto. Genetic algorithm application in solving crystallographic phase problem. *AIP Conference Proceedings*, 989:214–217, 2008.

[2] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

[3] E. D. Andersen, C. Roos, and T. Terlaky. On implementing a primal-dual interior-point method for conic quadratic optimization. *Mathematical Programming*, 95(2):249–277, 2003.

[4] A. Auger. *Theory of randomized search heuristics : Foundations and recent developments*. World Scientific & Inperial Colledge Press, 2011.

[5] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, volume 2 of *MPS/SIAM Series on Optimization*. SIAM, Philadelphia, PA, 2001.

[6] A. Ben-Tal and A. Nemirovski. On polyhedral approximations of the second-order cone. *Mathematics of Operations Research*, 26(2):193–205, 2001.

[7] A. Benediktovitch, I. Feranchuk, and A. Ulyanenkov. *Theoretical Concepts of X-Ray Nanoscale Analysis*. Springer, 2014.

[8] K. P. Bennett and O. L. Mangasarian. Robust linear programming separation of two linearly inseparable sets. *Optimization Methods and Software*, 1(1):23–34, 1992.

[9] Hande Y. Benson, Robert J. Vanderbei, and David F. Shanno. Interior-point methods for nonconvex nonlinear programming: Filter methods and merit functions. *Computational Optimization and Applications*, 23(2):457–272, 2002.

[10] S. Boyd and L. Vanderberge. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

[11] A. T. Bronger and L. M. Rice. Crystallographic refinement by simulated annealing: Methods and applications. 1997.

[12] B. D. Cullity. *Elements of X-ray difraction*. Addison-Wesley Publishing Company, Inc., 1956.

[13] E. de Klerk. *Aspects of Semidefinite Programming*. Kluwer Academic Publishers, Dordrecht, 2004.

[14] K. Doll, J. C. Schon, and M. Jansen. Structure prediction based on ab initio simulated annealing. 2008.

[15] F. Glineur. Computational experiments with a linear approximation of second order cone optimization. Technical Report 0001, Service de Mathématique et de Recherche Opérationnelle, Faculté Polytechnique de Mons, Mons, Belgium, 2000.

[16] F. Glineur and T. Terlaky. Conic formulation for l_p-norm optimization. *Journal of Optimization Theory and Applications*, 122(2):285–307, 2004.

[17] A. Guinier. Small angle scattering meeting: Summarizing remarks. *Journal of applied physics*, 30:601–603, 1959.

[18] B. Hajek. Cooling schedules for optimal annealing. *Mathematics of Operations Research*, 13(2):311–329, 1988.

[19] Kenneth D. M. Harrisa, Roy L. Johnston, and Benson M. Kariuki. The genetic algorithm: Foundations and applications in structure solution from powder diffraction data. *Acta Crystallographica*, A54:632–645, 1998.

[20] W. Helton and J. Nie. Semidefinite representation of convex sets. *Mathematical Programming*, 122(1):21–64, 2010.

[21] W. Helton and V. Vinnikov. Linear matrix inequality representation of sets. *Communications in Pure and Applied Mathematics*, 60(5):654–674, 2007.

[22] R. Hill. Theory of mechanical properties of fibre-strengthened materials: 1. inelastic behaviour. *Journal of the mechanics of Physics of Solids*, 12:213–218, 1964.

[23] F. Joda, N. Tahouni, and M. H. Panjeshahi. Application of genetic algorithms in design and optimisation of multi-stream platedfin heat exchangers. *The Canadian Journal of Chemical Engineering*, 91:870–881, 2013.

[24] P. G. Jones. Crystal structure determination: A critical view. *Chemical Society Reviews*, 13:157–172, 1984.

[25] D. Kalyanmoy. Genetic algorithm in search and optimization: The technique and applications. *Proceedings of International Workshop on Soft Computing and Intelligent Systems, (ISI, Calcutta, India)*, pages 58–87, 1998.

[26] B. M. Kariuki, S. A. Belmonte, M. I. McMahon, R. L. Jonston, K. D. M. Harris, and R. J. Nelmes. A new approach for indexing powder diffraction data based on whole-profile fitting and global optimization using a genetic algorithm. *Journal of Synchrotron Radiation*, 6:87–92, 1999.

[27] S. Kirkpatrick, Jr. C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[28] P. Krokhmal and P. Soberanis. Risk optimization with p-order conic constraints: A linear programming approach. *European Journal of Operational Research*, 301(3):653–671, 2010.

[29] Z. Liang, K. R. Shankar, K. Barefield, L. Zhang, C. Kramer, and B. Wang. Investigation of magnetically aligned carbon nanotube bucky papers/epoxy composites. In *Proceedings of SAMPE (48th ISSE)*, Long Beach, CA, 2003.

[30] G. W. Milton. *The Theory of Composites*. Cambridge University Press, Cambridge, UK, 2002.

[31] Y. E. Nesterov and A. Nemirovski. *Interior Point Polynomial Algorithms in Convex Programming*, volume 13 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994.

[32] Y. E. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for self-scaled cones. *Mathematics of Operations Research*, 22:1–42, 1997.

[33] Y. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on Optimization*, 8:324–364, 1998.

[34] Yurii E. Nesterov. Towards non-symmetric conic optimization. *Optimization Methods & Software*, 27(4–5):893–917, 2012.

[35] R. Pannetier, L. Bassas-Alsina, O. Rodriguez-Carvajal, and V. S. Caignaert. Prediction of crystal structures from crystal chemistry rules by simulated annealing. *Nature*, 346:343–345, 1990.

[36] V. K. Pecharsky and P. Y. Zavalij. *Fundamentals of Powder Diffraction and Structural Characterization of Materials*. Springer, 2009.

[37] M. Rothschild and J. Stiglitz. Increasing risk i: a definition. *Journal of Economic Theory*, 2(3):225–243, 1970.

[38] J. F. Sturm. Using sedumi 1.0x, a matlab toolbox for optimization over symmetric cones. *Manuscript*, 1998.

[39] T. Terlaky. On l_p programming. *European Journal of Operational Research*, 22(1):70–100, 1985.

[40] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 1953 edition, 1944.

[41] L. J. Walpole. On the overall elastic moduli of composite materials. *Journal of the mechanics of Physics of Solids*, 17:235–251, 1969.

[42] A.J.C. Wilson. *Elements of X-ray Crystallography*. Addison-Wesley Publishing Company, 1970.

[43] Guoliang Xue and Yinyu Ye. An efficient algorithm for minimizing a sum of p-norms. *SIAM Journal on Optimization*, 10(2):551–579, 2000.

[44] X. Zhu, R. Birringer, U. Herr, and H. Gleiter. X-ray diffraction studies of the structure of nanometer-sized crystalline materials. *Physical Review B*, 35(17):9085–9090, 1987.

[45] S. G. Zhukov, V. V. Chernyshev, E. V. Babaev, E. J. Sonneveld, , and H. Schenk. Application of simulated annealing approach for structure solution of molecular crystals from X-ray laboratory powder data. *Z. Kristallogr.*, 216:5–9, 2001.

[46] O. I. Zhupanska. The effect of orientational distribution of nanotubes on buckypaper nanocomposite mechanical properties. *Mechanics of Advanced Materials and Structures*, 20(1):1–10, 2002.