PARAMETRIC AND MULTIOBJECTIVE OPTIMIZATION WITH APPLICATIONS

PARAMETRIC AND MULTIOBJECTIVE OPTIMIZATION WITH APPLICATIONS IN FINANCE

By

Oleksandr Romanko, B.Sc., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2010) McMaster University
(Computing and Software) Hamilton, Ontario

TITLE:	Parametric and Multiobjective Optimization with Applications in Finance
AUTHOR:	Oleksandr Romanko, B.Sc., M.Sc.
SUPERVISORS:	Dr. Tamás Terlaky, Dr. Antoine Deza
NUMBER OF PAGES:	xx, 242

Abstract

In this thesis parametric analysis for conic quadratic optimization problems is studied. In parametric analysis, which is often referred to as parametric optimization or parametric programming, a perturbation parameter is introduced into the optimization problem, which means that the coefficients in the objective function of the problem and in the right-hand-side of the constraints are perturbed. First, we describe linear, convex quadratic and second order cone optimization problems and their parametric versions. Second, the theory for finding solutions of the parametric problems is developed. We also present algorithms for solving such problems. Third, we demonstrate how to use parametric optimization techniques to solve multiobjective optimization problems and compute Pareto efficient surfaces.

We implement our novel algorithm for bi-parametric quadratic optimization. It utilizes existing solvers to solve auxiliary problems. We present numerical results produced by our parametric optimization package on a number of practical financial and non-financial computational problems. In the latter we consider problems of drug design and beam intensity optimization for radiation therapy.

In the financial applications part, two risk management optimization models are developed or extended. These two models are a portfolio replication framework and a credit risk optimization framework. We describe applications of multiobjective optimization to existing financial models and novel models that we have developed. We solve a number of examples of financial multiobjective optimization problems using our parametric optimization algorithms.

Acknowledgments

This thesis was written under the guidance and with the help of my supervisor, Prof. Tamás Terlaky. His valuable advice and extensive knowledge of the area helped and motivated me in my work on this thesis. Prof. Antoine Deza became my co-supervisor when I was half-way through my studies. His supervision significantly expedited my skills and professional development. I am very grateful to Dr. Alireza Ghaffari-Hadigheh, Dr. Helmut Mausser, Dr. Alexander Kreinin and Dr. Ian Iscoe for their contribution to my thesis. My special thanks are to the members of the supervisory committee: Dr. Antoine Deza, Dr. Frantisek Franek, Dr. Tom Hurd and Dr. Tamás Terlaky.

It would not be possible to complete this thesis without support and help of all members of the Advanced Optimization Laboratory and the Department of Computing and Software. I am especially grateful to Olesya Peshko, Dr. Imre Pólik and Dr. Yuriy Zinchenko for their time and helpful suggestions. I would like to acknowledge McMaster University for the Sherman Graduate Scholarship and OGSST Scholarship, and Government of Ontario for the Ontario Graduate Scholarship. I am thankful to MITACS for two MITACS ACCELERATE Internships and a number of poster competition prizes that encouraged me in my work on the thesis. My special thanks are to Algorithmics Incorporated for help and inspiration during my internships there.

Finally, I appreciate the support of my parents, Serhiy and Vira Romanko, and friends, and thankful to them for their patience and understanding.

Contents

Abbreviations

CAPM:	Capital Asset Pricing Model
cdf:	cumulative distribution function
CLO:	conic linear optimiztion
CLT:	central limit theorem
CML:	capital market line
CP:	counterparty
CQO:	conic quadratic optimiztion
ES, CVaR:	expected shortfall (conditional value-at-risk)
EUD:	equivalent uniform dose
IMRT:	intensity modulated radiation therapy
IPC:	interior point condition
IPM:	interior point method
LASSO:	least absolute shrinkage and selection operator
LLN:	law of large numbers
LO:	linear optimization
MC:	Monte Carlo
MV:	mean-variance

pdf:	probability density function
QCAR:	quantitative composition-activity relationship
QCQO:	quadratically constrained quadratic optimization
QO:	(convex) quadratic optimization
SDO:	semidefinite optimization
SOCO:	second-order cone optimization
VaR:	value-at-risk

Notation and Symbols

Basic Objects

$A, B, \ldots$	matrices
A_{ij}	the j^{th} element of the i^{th} row of A
$x, y, \ldots$	column vectors
x_i	the i^{th} component of x
x^k	the k^{th} vector in a list
$i, j, \ldots$	indices
$x_{i:j}$	the vector $(x_i, \ldots, x_j)^T$
$\mathbf{1}, \mathbf{1}_n$	the (n-dimensional) vector of ones
$f, g, \ldots$	functions
$\mathcal{K}$	convex (usually quadratic) cone
$\mathcal{C}, \Omega, \ldots$	sets
$R, Y, \ldots$	random variables

Operators, Functions and Sets

$\mathbb{R}^n$	the real n-dimensional vector space
$\mathbb{R}^n_+$	the set of n-dimensional vectors with nonnegative components
$\mathcal{K}^*$	the dual cone of $\mathcal{K}$
$\operatorname{int}\mathcal{K}$	the interior of $\mathcal{K}$
$\operatorname{bd}\mathcal{K}$	the boundary of $\mathcal{K}$
$\|x\|_2$	Euclidian norm, $\|x\|_2 = \sqrt{x^T x} = \sqrt{x_1^2 + \ldots + x_n^2}$
$\|x\|_1$	ℓ_1 norm, $\|x\|_1 = \sum_{i=1}^n \|x_i\|$
$\|x\|_\infty$	infinity norm, $\|x\|_\infty = \max\{\|x_1\|, \ldots, \|x_n\|\}$
$\|x\|_p$	p norm, $\|x\|_p = (\sum_{i=1}^n \|x_i\|^p)^{1/p}$
$\operatorname{card}(x)$	vector cardinality, $\operatorname{card}(x) = \|x\|_0 = \sum_{i=1}^n x_i^0$ with $0^0 := 0$
$\|\mathcal{C}\|$	set cardinality, number of elements in a finite set $\mathcal{C}$
$\mathbb{E}[\cdot]$	expected value
$\operatorname{Var}[\cdot]$	variance
$\operatorname{Cov}[\cdot]$	covariance

Relations

$x \geq 0$	the components of x are all nonnegative
$A \succeq 0$	A is symmetric positive semidefinite
$A \succ 0$	A is symmetric positive definite
$x \geq_{\mathcal{K}} 0$	$x \in \mathcal{K}$
$x >_{\mathcal{K}} 0$	$x \in \operatorname{int}\mathcal{K}$

List of Figures

List of Tables

Chapter 1

Introduction

Optimization is a discipline used for searching extremum of a function and generally refers to mathematical problems where the goal is to minimize or maximize an objective function subject to some constraints. Depending on the nature and the form of the objective function and the constraints, continuous optimization problems are classified to be linear, quadratic, conic and general nonlinear. Correspondingly, we distinguish the research areas of linear optimization, quadratic optimization, second order conic optimization, etc.

1.1 Conic Quadratic Optimization Problems

Linear optimization (LO) is a highly successful operations research model. Therefore, it was natural to generalize the LO model to handle more general nonlinear relationships. However, this gives rise to many difficulties such as lack of strong duality, possible non-convexity and consequently problems with global versus local optimums, lack of efficient algorithms and software, etc.

In the recent decade, a new class of convex optimization models that deals with the problem of minimizing a linear function subject to an affine set intersected with a convex cone has appeared. It is known as conic optimization. Although the conic optimization model seems to be restrictive, any convex optimization problem can be cast as a conic optimization model and there are effi-

cient solution algorithms for many classes of conic models such as conic quadratic optimization (CQO) and conic linear optimization (CLO). While it sounds counterintuitive, CQO is a sub-class of CLO. Conic optimization has many interesting applications in engineering, image processing, finance, economics, combinatorial optimization, etc.

1.1.1 Formulation of Conic Quadratic Optimization Problems

Conic linear optimization (CLO) problems can be mathematically formulated as:

$$\min_x \{c^T x : Ax = b,\ x \in \mathcal{K}\}, \tag{1.1.1}$$

where $\mathcal{K} \in \mathbb{R}^n$ is a closed, convex, pointed and solid cone, $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = m$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ are fixed data and $x \in \mathbb{R}^n$ is an unknown vector. Often $x \in \mathcal{K}$ is also denoted as $x \succeq_{\mathcal{K}} 0$. Moreover, $x \succeq_{\mathcal{K}} y$ means $x - y \succeq_{\mathcal{K}} 0$.

Definition 1.1.1 (Cone) *A set $\mathcal{C}$ is a (linear) cone if $\forall x \in \mathcal{C}$ and $\alpha \geq 0$, $\alpha x \in \mathcal{C}$.*

Definition 1.1.2 (Convex cone) *A subset $\mathcal{K}$ of $\mathbb{R}^n$ is a convex cone if and only if $\alpha_1 x^1 + \alpha_2 x^2 \in \mathcal{K}$ for any $\alpha_1, \alpha_2 \geq 0$ and $x^1, x^2 \in \mathcal{K}$.*

Observe that every cone includes the null vector. A pointed convex cone contains no line. Solid cone has non-empty interior. The set of all interior points of $\mathcal{K}$ is denoted by $\text{int}(\mathcal{K})$.

Examples of pointed convex closed cones include:

- the nonnegative orthant:

$$\mathbb{R}^n_+ = \mathcal{K}_\ell = \{x \in \mathbb{R}^n : x \geq 0\},$$

- the quadratic cone (also know as Lorentz cone, second order cone or ice-cream cone:

$$L^n = \mathcal{K}_q = \{x \in \mathbb{R}^n : x_1 \geq \|x_{2:n}\|_2\},$$

- the semidefinite cone:

$$S_+^n = \mathcal{K}_s = \{X \in \mathbb{R}^{n \times n} : X = X^T,\ X \succeq 0\},$$

- any linear transformation and finite direct product of such cones.

Definition 1.1.3 (Cartesian product) *If $\mathcal{A} \subseteq \mathbb{R}^j$ and $\mathcal{B} \subseteq \mathbb{R}^k$, then*

$$\mathcal{A} \times \mathcal{B} = \{(x, y) : x \in \mathcal{A},\ y \in \mathcal{B}\}$$

is their Cartesian product (direct product).

The definition of the dual cone arises from the question: for which $s \in \mathbb{R}^n$ the implication $x^1 \geq_{\mathcal{K}} x^2 \Rightarrow s^T x^1 \geq s^T s^2$ always holds. Obviously, this holds if and only if $x \geq_{\mathcal{K}} 0$ implies $s^T x \geq 0$, or if $s \in \mathcal{K}^* = \{s \in \mathbb{R}^n : s^T x \geq 0,\ \forall x \in \mathcal{K}\}$.

Definition 1.1.4 (Dual cone) *The dual cone $\mathcal{K}^*$ of a convex cone $\mathcal{K} \in \mathbb{R}^n$ is*

$$\mathcal{K}^* = \{s \in \mathbb{R}^n : s^T x \geq 0,\ \forall x \in \mathcal{K}\}.$$

Theorem 1.1.1 *If $\mathcal{K}$ is a closed, convex, pointed cone with nonempty interior, then so is the dual cone $\mathcal{K}^*$, and then the duality is symmetric: $(\mathcal{K}^*)^* = \mathcal{K}$.*

Each of the three "standard" cones $\mathcal{K}_\ell$, $\mathcal{K}_q$ and $\mathcal{K}_s$ are closed, convex and pointed cones with nonempty interior. Moreover, each of these cones are self-dual, which means that the dual cone $\mathcal{K}^*$ is equal to the original cone $\mathcal{K}$. The same holds for any (finite) direct product of such cones.

Now we present the dual conic linear optimization problem:

$$\max_{y,\,s} \{b^T y : A^T y + s = c,\ s \in \mathcal{K}^*\}. \tag{1.1.2}$$

If x is feasible for (1.1.1) and (y, s) is feasible for (1.1.2), then the *weak duality* property holds

$$c^T x - b^T y = x^T s \geq 0. \tag{1.1.3}$$

The *strong duality* property $c^T x = b^T y$ does not always hold for CLO problems. A sufficient condition for strong duality is the primal-dual Slater condition, which requires the existence of a feasible solution pair x and (y, s) for (1.1.1) and (1.1.2) such that $x \in \operatorname{int} \mathcal{K}$ and $s \in \operatorname{int} \mathcal{K}^*$. In this case, the primal-dual optimal set of solutions (x, y, s) is

$$\begin{array}{rcll} Ax & = & b, & x \in \mathcal{K}, \\ A^T y + s & = & c, & s \in \mathcal{K}^*, \\ x^T s & = & 0. & \end{array} \tag{1.1.4}$$

System (1.1.4) is known as the *optimality conditions*.

Conic quadratic optimization (CQO) is the problem of minimizing a linear objective function subject to the intersection of an affine set and the direct product of quadratic cones. CQO is the sub-class of CLO and, consequently, CQO problems are expressed in the form of (1.1.1).

More information on CLO and CQO problems, their properties and duality results can be found in [8]. The CQO problem subclasses described in the following sections include linear optimization (LO), convex quadratic optimization (QO), quadratically constrained quadratic optimization (QCQO) and second-order conic optimization (SOCO). CLO, among others, includes CQO and semidefinite optimization (SDO). As we see in Section 1.1.5, in all these cases CLO problems can be solved efficiently by Interior Point Methods (IPMs).

1.1.2 Linear Optimization

The best known and most widely used form of optimization models is Linear Optimization (LO), where the objective function and all the constraints are linear functions of the variables. Although this form seems to be restrictive, it is widely used in practice, especially in business applications, because many practical problems can be formulated with linear objectives (such as profits) and linear constraints (budgets, capacities, etc.). Many practical problems – especially for large companies – involve hundreds of thousands to millions of variables and constraints, and LO problems of this size can be solved quickly and reliably with modern software.

LO is the most extensively studied branch in the optimization field. History of linear optimization started in the 40's of the 20th century and has gained wide attention by the scientific community in the 50's after the development of the Simplex method by George Dantzig [34]. The Simplex algorithm and its extensions were thoroughly studied since then, and did not have practical competitors until the discovery of Interior Point Methods (IPMs) in the middle of the 80's. The milestone work of Karmarkar [82] in 1984 started the era of IPMs, that usually outperform simplex algorithms when solving large-scale problems.

The main conceptual feature that differentiates IPMs from the Simplex method is how they search for an optimal solution. In simplex methods the optimal solution is searched by moving from one vertex of the feasible region to another until an optimal (minimum or maximum) solution is found, while in IPMs the problem is solved by following a path inside the feasible region of the problem that leads to optimality. IPMs allow solving large sparse optimization problems efficiently in polynomial time. Reference [106] provides a comprehensive description of simplex methods and [126] gives the theoretical and algorithmic background of IPMs for linear optimization.

As we already know, we can get the LO problem by restricting the CQO problem to have the positive orthant as its cone. The primal LO problem is:

$$(LP) \qquad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0, \end{array} \tag{1.1.5}$$

where $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = m$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ are fixed data and $x \in \mathbb{R}^n$ is an unknown vector.

Its dual form is:

$$(LD) \qquad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + s = c \\ & s \geq 0. \end{array} \tag{1.1.6}$$

1.1.3 Quadratic Optimization

Convex Quadratic Optimization (QO) problems, where the objective function is convex quadratic while the constraints remain linear, gained its importance among business community after Markowitz [95, 96] used it for conducting mean-variance analysis of investment portfolios. QO problems appear naturally when we want to minimize a variation (or variance) of some quantity. In general, QO is a natural extension of LO and most of the solution methods developed for LO were extended to QO as well.

A primal convex QO problem is defined as:

$$(QP) \qquad \begin{array}{ll} \min & c^T x + \frac{1}{2} x^T Q x \\ \text{s.t.} & Ax = b \\ & x \geq 0, \end{array} \tag{1.1.7}$$

where $Q \in \mathbb{R}^{n \times n}$ is a symmetric positive semidefinite matrix, $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = m$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ are fixed data and $x \in \mathbb{R}^n$ is an unknown vector.

The Wolfe Dual of (QP) is given by

$$(QD) \qquad \begin{array}{ll} \max & b^T y - \frac{1}{2} u^T Q u \\ \text{s.t.} & A^T y + s - Qu = c \\ & s \geq 0, \end{array} \tag{1.1.8}$$

where $s, u \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are unknown vectors.

We provide more details about QO problems to facilitate the discussion of their parametric counterparts in Chapter 2. All the concepts described below are valid too for LO problems where $Q = 0$.

The feasible regions of (QP) and (QD) are denoted by

$$\begin{aligned} \mathcal{QP} &= \{x : Ax = b,\ x \geq 0\}, \\ \mathcal{QD} &= \{(u, y, s) : A^T y + s - Qu = c,\ s, u \geq 0\}, \end{aligned}$$

and their associated optimal solution sets are denoted by $\mathcal{QP}^*$ and $\mathcal{QD}^*$, respectively. It is known that for any optimal solution of (QP) and (QD) we have $Qx = Qu$ and $x^T s = 0$, e.g., see Dorn [40]. It is also known from [40] that there

are optimal solutions with $x = u$. Since we are only interested in the solutions where $x = u$, therefore, u will be replaced by x in the dual problem. The duality gap $c^T x + x^T Q x - b^T y = x^T s$ being zero is equivalent to $x_i s_i = 0$ for all $i \in \{1, 2, \ldots, n\}$. This property of the nonnegative variables x and s is called the *complementarity property*.

Solving primal problem (QP) or dual problem (QD) is equivalent to solving the following system, which represents the Karush-Kuhn-Tucker (KKT) *optimality conditions* [151]:

$$\begin{array}{rcll} Ax - b & = & 0, & x \geq 0, \\ A^T y + s - Qx - c & = & 0, & s \geq 0, \\ x_i s_i & = & 0, & \forall i, \end{array} \tag{1.1.9}$$

where the first line is the primal feasibility, the second line is the dual feasibility, and the last line is the complementarity condition. The complementarity condition can be rewritten as $xs = 0$, where xs denotes the componentwise product of the vectors x and s. System (1.1.9) is referred to as the *optimality conditions*.

For LO the Goldman-Tucker Theorem states that there exists a *strictly complementary optimal solution* (x, y, s) if both the primal and dual problems are feasible. The feasible primal-dual pair (x, y, s) is strictly complementary if $x_i s_i = 0$ and $x_i + s_i > 0$ for all $i = 1, \ldots, n$. Equivalently, strict complementarity can be characterized by $xs = 0$ and $x + s > 0$. A strictly complementary solution partitions the index set in two subsets, the sets where $x_i > 0$, or $s_i > 0$, respectively. Having strict complementarity is important for IPM-based parametric analysis.

Unlike in LO, where strictly complementary optimal solution always exists, for QO the existence of such solution is not ensured. Instead, a *maximally complementary solution* can be found. A pair of optimal solutions (x, y, s) for the QO problem is maximally complementary if it maximizes the number of nonzeros in the vector $x + s$ over all optimal solution pairs. As we see in Section 2.1, this leads to tri-partition of the optimal solution set in parametric QO.

1.1.4 Second-Order Conic Optimization

We have seen that a linear constraint qualifies as a cone constraint; the next step is the so-called second-order cone, also referred to as the Lorenz cone, quadratic cone or ice-cream cone. QO is also a subclass of Second-Order Conic Optimization (SOCO) problems.

In SOCO problems the variables are restricted to lie in the Lorentz cone leading to the following formulation:

$$\begin{array}{ll} & \min \quad c^T x \\ (SOCP) & \text{s.t.} \quad Ax = b, \\ & \qquad x_1^i \geq \|x_{2:n_i}^i\|_2, \; i = 1, \ldots, I, \end{array} \tag{1.1.10}$$

where $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = m$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ are fixed data and $x = (x_1^1, \ldots, x_{n_1}^1, x_1^2, \ldots, x_{n_2}^2, x_1^3, \ldots, x_{n_I}^I)^T \in \mathbb{R}^n$ (with $n = \sum_{i=1}^I n_i$) is an unknown vector.

The dual SOCO problem is:

$$\begin{array}{ll} & \max \quad b^T y \\ (SOCD) & \text{s.t.} \quad A^T y + s = c, \\ & \qquad s_1^i \geq \|s_{2:n_i}^i\|_2, \; i = 1, \ldots, I, \end{array} \tag{1.1.11}$$

where $y \in \mathbb{R}^m$ and $s = (s_1^1, \ldots, s_{n_1}^1, s_1^2, \ldots, s_{n_2}^2, s_1^3, \ldots, s_{n_I}^I)^T \in \mathbb{R}^n$ are unknown vectors.

As $(x_1^1, \ldots, x_{n_1}^1)^T \in \mathcal{K}_q^1$, $(x_2^1, \ldots, x_{n_2}^2)^T \in \mathcal{K}_q^2$, ..., $(x_1^I, \ldots, x_{n_I}^I)^T \in \mathcal{K}_q^I$ and $\mathcal{K} = \mathcal{K}_q^1 \times \mathcal{K}_q^2 \times \ldots \times \mathcal{K}_q^I$ we can also rewrite problem (1.1.10) in its shorter form

$$\begin{array}{ll} & \min \quad c^T x \\ (SOCP) & \text{s.t.} \quad Ax = b, \\ & \qquad x \in \mathcal{K}. \end{array} \tag{1.1.12}$$

In the remainder of the thesis, cone $\mathcal{K}$ denotes the quadratic cone (direct product of linear cones $\mathcal{K}_\ell$ and quadratic cones $\mathcal{K}_q$), unless otherwise specified.

In addition to LO and QO problems, SOCO also includes quadratically constrained quadratic optimization (QCQO). Details about the QCQO problem formulation and its transformation to SOCO formulation can be found in [91].

1.1.5 Interior Point Methods for Quadratic Optimization

As it was mentioned before, the seminal work of Karmarkar [82] resulted in the discovery of IPMs for solving LO problems. IPMs enjoy worst-case polynomial complexity and excellent practical performance on large-scale LO problems.

The field of IPMs received more attention after Nesterov and Nemirovskii [107] unified the theory of IPMs into the general framework and extended it to general convex optimization. Their work made it possible to develop IPMs for solving much larger classes of optimization problems including QO, SOCO and SDO.

Primal-dual IPMs are iterative algorithms that aim to find a solution satisfying the optimality conditions. For instance, optimality conditions for the QO case are expressed by the system (1.1.9). Here, we briefly describe feasible IPMs for QO (IPMs for the CQO are the generalizations of those). IPMs generate a sequence of iterates (x^k, y^k, s^k), $k = 0, 1, 2, \ldots$ that satisfy the strict positivity (interior point) condition $x^k > 0$ and $s^k > 0$, but feasibility (for infeasible IPMs) and optimality are reached as k goes to infinity. Feasible IPMs produce a sequence of iterates where the following *interior point condition* (IPC) holds for every iterate (x, y, s)

$$\begin{aligned} Ax &= b, \quad x > 0, \\ A^T y + s - Qx &= c, \quad s > 0. \end{aligned} \tag{1.1.13}$$

We perturb the complementarity condition (1.1.9) as

$$\begin{aligned} Ax &= b, \ x > 0, \\ A^T y + s - Qx &= c, \ s > 0, \\ Xs &= \mu e, \end{aligned} \tag{1.1.14}$$

where $\mu > 0$, $e = (1, \ldots, 1)^T$ and $X = \mathrm{diag}(x_1, \ldots, x_n)$ is the diagonal matrix with vector x forming the diagonal. It is obvious that the last nonlinear equation in (1.1.14) becomes the complementarity condition for $\mu = 0$.

A desired property of system (1.1.14) is the uniqueness of its solution for each $\mu > 0$. The following theorem [66] provides the conditions when uniqueness holds.

Theorem 1.1.2 *System (1.1.14) has a unique solution for each $\mu > 0$ if and only if* $\text{rank}(A) = m$ *and the IPC holds for some point.*

When μ is running through all positive numbers, the set of unique solutions $(x(\mu), y(\mu), s(\mu))$ of (1.1.14) define the so-called primal-dual central path. The sets $\{x(\mu) \mid \mu > 0\}$ and $\{(y(\mu), s(\mu)) \mid \mu > 0\}$ are called the *primal central path* and the *dual central path* respectively.

One iteration of primal-dual IPMs consists of taking a Newton step applied to the central path equations (1.1.14) for a given μ. The central path stays in the interior of the feasible region and the algorithm approximately follows it towards optimality. For $\mu \to 0$ the set of points $(x(\mu), y(\mu), s(\mu))$ gives us a *maximally complementary* (strictly complementary in LO case and maximally complementary in SOCO case) optimal solution of (QP) and (QD).

Newton's method is used to solve the system (1.1.14) iteratively. At each step we need to compute the direction $(\triangle x, \triangle y, \triangle s)$. A new point in the computed direction $(x + \triangle x, y + \triangle y, s + \triangle s)$ should satisfy

$$\begin{aligned} A\triangle x &= 0, \\ A^T\triangle y + \triangle s - Q\triangle x &= 0, \\ x\triangle s + s\triangle x + \triangle x\triangle s &= \mu e - xs. \end{aligned} \tag{1.1.15}$$

System (1.1.15) is non-linear. Consequently, the Newton step is obtained by dropping the non-linear term that gives the linearized Newton system

$$\begin{aligned} A\triangle x &= 0, \\ A^T\triangle y + \triangle s - Q\triangle x &= 0, \\ x\triangle s + s\triangle x &= \mu e - xs. \end{aligned} \tag{1.1.16}$$

The linear system (1.1.16) is referred to as the *primal-dual Newton system.* It has $2n + m$ equations and $2n + m$ unknowns. The system has a unique solution if $\text{rank}(A) = m$.

After solving the Newton system (1.1.16), we have the search direction $(\Delta x, \Delta y, \Delta s)$. This Newton search direction is computed assuming that the step length α is equal to one. But taking such a step can lead to infeasibility of the

solution as $(x+\Delta x, y+\Delta y, s+\Delta s)$ might be infeasible. Our goal is to keep strict feasibility, therefore we want to find such an α that the next iteration point is strictly feasible, i.e.,

$$(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + \alpha(\Delta x^k, \Delta y^k, \Delta s^k),$$

with $x^{k+1} > 0$ and $s^{k+1} > 0$. This can be done via line search.

Based on the LO case derivations above, here we present a prototype Primal-Dual Path-Following IPM algorithm (see Algorithm 1.1) for LO and CQO problems. We suppose that the IPC is satisfied. Alternatively, by using the homogeneous embedding technique (see [126] for details) we can always construct a LO or QO problem in a way that the IPC holds. After that, we apply Newton's method to the central path equations (1.1.14) to get the search direction. The determined step length ensures that the iterate remains in the interior of the feasible set. IPMs stop when the complementarity gap is reduced below some predetermined tolerance level or when infeasibility is detected.

One issue that we have not discussed so far is a strategy for reducing the parameter μ. We want to follow the central path approximately (as Newton system is solved disregarding the nonlinear terms) and so we define a proximity function $\Psi(xs, \mu)$ to measure the distance of the current point from the central path. A proximity parameter $\delta > 0$ defines the bound for the proximity function which takes the value of zero if the point is on the central path and approaches infinity if the point approaches the boundary of the nonnegative orthant. There are many proximity measures defined in the literature (see, e.g., [126] for more details), for instance, self-regular proximity functions.

1.2 Parametric Optimization

Uncertainty is a very important factor in many optimization models. For instance, when uncertainty arises in some of the parameters, the optimal solution of a deterministic model can become infeasible. Therefore, consideration of un-

Algorithm 1.1: A Prototype Algorithm for Interior Point Method.

input:
a proximity parameter $\delta > 0$;
an accuracy parameter $\epsilon > 0$;
an update parameter $0 < \theta < 1$;
$\mu_0 = 1$, $k = 0$;
(x^0, y^0, s^0) satisfying $x^0 > 0$, $s^0 > 0$ and $\Psi(x^0 s^0, \mu^0) \leq \delta$;
begin
while $(x^k)^T s^k \geq \epsilon$ **do**
begin
$\mu^k = (1-\theta)(\frac{(x^k)^T s^k}{n})$;
while $\Psi(x^k s^k, \mu^k) \geq \delta$ **do**
begin
solve the Newton system to find $(\triangle x^k, \triangle y^k, \triangle s^k)$;
determine the step size α to get strictly interior point;
$x^k = x^k + \alpha \triangle x^k$, $y^k = y^k + \alpha \triangle y^k$, $s^k = s^k + \alpha \triangle s^k$;
end
$x^{k+1} = x^k$, $y^{k+1} = y^k$, $s^{k+1} = s^k$, $k = k + 1$;
end
end

certainty becomes of great importance in order to preserve feasibility and reliability of the results. One way to incorporate uncertainty into optimization models is using *sensitivity analysis* and *parametric optimization* techniques. Other approaches to deal with uncertainty include *robust optimization* and *stochastic optimization*.

In addition to the uncertainty factor, in most of the practical applications we are interested not only in the value of an optimal solution of an optimization problem, but also in its sensitivity. In other words, it is necessary to know how sensitive the solution is to data perturbations. Knowing sensitivity of the solution allows adjusting the constraints (such as resource constraints or budget constraints) or the coefficients of the objective function (such as individual preferences or parameters of a production process) to meet the modeling objectives

in a better way and to get a "better" solution to the problem.

Similarly to the classical sensitivity analysis, where only one model parameter is perturbed, in *parametric optimization* a number of model parameters, which are held constant in the original deterministic problem, are assumed to be uncertain or variable. Thus, mathematical models in general and optimization problems in particular often involve unknown or uncertain parameters and the goal of parametric optimization is to calculate the optimal solution for all relevant values of these unknown parameters. Discretization approaches are non-rigorous since there is no guarantee for optimality between the mesh points. Therefore, algorithms for parametric optimization typically divide the parameter space into regions of optimality; for each region infeasibility is established or an optimal solution is given as a smooth function of the parameters for this region. For one parameter, the transition between the solutions is called a breakpoint or a transition point and parametric optimization algorithms identify these breakpoints.

As we see both sensitivity analysis and parametric optimization (we will use the terms *sensitivity analysis*, *parametric analysis* and *parametric optimization* interchangeably thereafter) are the techniques to determine how the optimal objective function value varies with the change in one or more coefficients of the objective function or the problem constraints. Mathematically speaking, in parametric analysis a parameter λ is introduced into the original optimization problem transforming it to the parametric one:

$$\begin{array}{lll} \min & f(x) & \\ \text{s.t.} & g_i(x) \leq 0, \quad \forall\, i & \end{array} \quad \Longrightarrow \quad \begin{array}{lll} \phi(\lambda) = & \min & f(x,\lambda) \\ & \text{s.t.} & g_i(x,\lambda) \leq 0, \quad \forall\, i. \end{array}$$

The parameter λ can be a scalar or multidimensional vector. Parametric optimization allows determining how the optimal objective function value $\phi(\lambda)$ (also called *optimal value function*) and an optimal solution $x^*(\lambda)$ varies with the change in one or more coefficients (parameters) of the objective function or right-hand side values of the constraints. Consequently, it allows us obtaining optimal solutions as a function of the parameters λ appearing in the optimization problem. Single-parameter problems (when λ is scalar) yield paths of solution points,

two parameter problems yield surfaces, etc. The main limitation of most existing parametric optimization methods is that they can only be applied to problems with a single uncertain parameter or several uncertain parameters varying in a single direction.

Parametric optimization is a relatively mature field with many contributions, e.g., see [106, 12, 123, 144] for reviews. With an exception of a few algorithms for linear or purely integer programs the available literature assumes that parameters only influence the objective function and/or the right-hand side of the constraints. Extension to the general case where a parameter can simultaneously affect the non-constant part of constraints, the right-hand side and the objective function is non-trivial because the resulting problems do not have the nice structure enjoyed by the special cases. Algorithms dealing with the general case are desired as well, but not well studied.

One of the goals of this thesis is to outline the possible use of the interior point methods framework to conducting parametric analysis of CQO problems. Developing and extending methodologies that allow finding an optimal solution vector $x^*(\lambda)$ and the optimal value function $\phi(\lambda)$ without discretization of the parameter space Λ and without solving the optimization problem at every discretization point is one of our primary targets.

1.3 Multiobjective Optimization

Multicriteria decision making or *multicriteria analysis* is a complex process of finding the best compromise among alternatives. A decision maker first describes the problem based on relevant assumptions about the real world problem. After that, alternative decisions are generated and evaluated. Optimization serves as a tool for solving multicriteria analysis problems when those problems are formulated as *multiobjective optimization* problems.

Let x be an n-dimensional vector of *decision variables*. The multiobjective optimization problem, where the goal is to optimize a number of possibly

conflicting objectives simultaneously, is formulated as:

$$\begin{array}{ll} \min & \{f_1(x), f_2(x), \ldots, f_k(x)\} \\ \text{s.t.} & x \in \Omega, \end{array} \tag{1.3.1}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, k$ are (possibly) conflicting objectives and $\Omega \subseteq \mathbb{R}^n$ is a feasible region. Each of the functions f_i represent an attribute or a *decision criterion* that serves the base for the decision making process. Similarly to single-objective optimization problems, depending on the nature and the form of the objective functions and the constraints, continuous multiobjective optimization problems are classified to linear, quadratic, conic and general nonlinear.

Multiobjective optimization is a subclass of *vector optimization*, where the vector-valued objective function $f_0 = \{f_1(x), f_2(x), \ldots, f_k(x)\}$ is optimized with respect to a proper convex cone $\mathcal{C}$ which defines preferences. When a vector optimization problem involves the cone $\mathcal{C} = \mathbb{R}_+$, it is known as a *multicriteria* or *multiobjective optimization* problem. Theoretical background and solution techniques for multiobjective optimization are discussed in Section 3.1.

We only look at some classes of convex multiobjective optimization problems, namely LO, QO and briefly at SOCO. In this case, all the functions appearing in the objective and the constraints of the optimization problems are conic quadratic. Most of real-life optimization problems are multiobjective in their nature and in many cases those can be formulated as multiobjective LO, QO or SOCO problems. Classical meaning of the word "optimization" refers to single-objective optimization, for instance, LO, QO and SOCO problems we described in Section 1.1 are single-objective convex optimization problems.

There are many relationships between multiobjective optimization and parametric optimization that is used to solve such problems, and in Section 3.2 we highlight those. The "solution" of a multiobjective problem is the set of Pareto efficient points, known in the literature as *Pareto efficient frontier* or *Pareto front*. Pareto points can be obtained by scalarization techniques that transform multiobjective optimization problem into series of single-objective problems. We can formulate that series of problems as parametric optimization problems and

compute its efficient solution set numerically. In Chapter 3, we present a methodology that allows not only tracing the Pareto efficient frontier without discretization of the objective space and without solving the corresponding optimization problem at each discretization point, but also identifying a structure of the frontier using parametric optimization. The main idea of Chapter 3 and one of the highlights of this thesis is that we can solve multiobjective optimization problems using parametric optimization techniques by systematically generating the Pareto front.

The main motivation for exploring multiobjective optimization is that it has numerous applications. Multiobjective optimization problems arise in many areas including engineering, where a typical goal may be to maximize vehicle speed and maximize its safety. For formulations of multiobjective optimization problems appearing in engineering we refer the reader to consult vast literature on multi-disciplinary design [3]. In finance commonly used conflicting objectives are maximizing profit and minimizing risk, while problems from environmental economics may involve maximizing profit and minimizing environmental impact. Examples described in this thesis, see Chapter 7, are mostly financial optimization problems from the area of risk management and portfolio selection. One of the multiobjective optimization problems in health care is to kill tumor and spare healthy tissues. Health care applications, discussed in Chapter 5, include intensity modulated radiation therapy (IMRT) planning for cancer treatment and optimal drug design. In Chapters 5 and 7 we not only present applications of multiobjective optimization, but also numerically solve a number of examples using our parametric optimization techniques.

1.4 Optimization in Finance

Optimization is one of the primary techniques for financial decision making. Wide use of optimization techniques in finance includes such classes of problems as *portfolio selection*, *risk management*, regression problems, pricing and hedging

of derivatives, and asset liability management. General overview of *financial optimization* or *optimization in finance* can be found in [30] and [157].

Linear optimization models are widely used in finance [94] due to availability of intuitive and relatively simple formulations and computational attractiveness. The class of quadratic optimization problems gained its importance among business community after Markowitz [95, 96] used it for conducting mean-variance analysis of investment portfolios. Even though second-order cone and semidefinite optimization problems are relatively new to the financial community, many financial problems can be solved using these techniques [30]. Robust portfolio optimization [47] is one of the most important among those. Convex and non-convex nonlinear optimization are the most complex continuous optimization techniques used in finance. Section 6.3 includes a number of novel nonlinear formulations for portfolio credit risk optimization that we have developed. Integer optimization is required to solve financial problems involving value-at-risk (VaR) minimization, cardinality constraints or fixed-plus-linear transaction cost. As most of financial optimization models need to deal with data uncertainty, robust optimization [7] and stochastic optimization [155] are also among the tools used by quantitative analysts. Another challenge that we would like to emphasize is that financial problems are usually large-scale and require solutions in near real-time.

Risk management is a systematic approach for minimizing exposure to risk. Usually, risks should be measured and managed across diverse instruments, geographies and risk types. Financial risks are classified to

- *market risk* – movement of an entire market;
- *volatility risk* – market volatility, influencing prices;
- *currency risk* – risk from international exposure;
- *credit risk* – risk that an obligor may default;
- *operational risk* – impact of operational events;
- *liquidity risk* – difficulty of selling an asset;
- company-specific risks, etc.

Risks are all related, so, risk management should be integrated, or at least interconnected. Investors and companies need to identify which risks to accept and which risks to protect against. Organizations and individuals need to analyze exposure to risk in order to determine how to optimize it. Optimization helps to compose diversified portfolios where different financial instruments offset or cancel each others' risks.

Many financial models involve maximizing a performance measure or minimizing risk associated with a financial decision. As performance usually conflicts with risk, trade-offs between performance measures and risks need to be identified and explored. In order to optimize risk, quantitative risk measures are required. Obviously, quantitative performance measures are needed as well.

Let us consider a typical risk management optimization problem. We denote by x_i a proportion of total funds invested in asset i, and by r expected return vector for different assets or some other asset performance indicator. Typical financial risk management optimization problem can be formulated as:

$$\begin{array}{ll} \max\limits_{x} & r^T x \\ \text{s.t.} & g(x) \leq \gamma \\ & \mathbf{1}^T x = 1 \\ & x \geq 0. \end{array} \tag{1.4.1}$$

In formulation (1.4.1), a performance measure $r^T x$ (i.e., expected investment return) is maximized subject to constraint $g(x) \leq \gamma$ that particular risk measure g does not exceed a prescribed amount γ and other operating constraints $\mathbf{1}^T x = 1, x \geq 0$ ($\mathbf{1}$ is the vector of 1's).

A closely related concept to risk management is *portfolio selection* or *portfolio optimization*, which deals with the problem of selecting an efficient portfolio of financial instruments. In *portfolio optimization*, the goal of investors is to obtain optimal returns in all market environments when risk is involved in every investment, borrowing, lending and project planning activity. From the multicriteria analysis point of view, investors need to determine what fraction of their wealth to invest in which asset in order to maximize the total return and mini-

mize the total risk of their portfolio. There are a number of risk measures used for quantitative evaluation of portfolio risk including variance, portfolio beta, value-at-risk (VaR) and expected shortfall (ES) among others, see Sections 6.1 and 6.3. In addition to risk measures, there are portfolio performance indicators – expected market return, expected credit loss, price earnings ratio, etc.

One of the most famous portfolio management models that involve a risk-return trade-off is the mean-variance portfolio optimization problem introduced by Markowitz [95], see Section 6.1 for more detail. The conflicting objectives in the Markowitz model are minimizing portfolio variance (risk) and maximizing expected return. Using variance to measure portfolio's risk, as proposed by Markowitz, is central to finance in both a theoretical and a practical sense. However, variance, which is expressed as a quadratic function of asset weights in a portfolio, is not the only risk measure that is used in practice. Multiobjective optimization is a natural tool for portfolio selection models as those involve minimizing one or several risk measures, and maximizing (for the return) or minimizing (for the losses) a number of portfolio performance indicators. Mean-variance optimization problem is one example of multiobjective problems in finance. We describe a number of variants of multiobjective portfolio optimization and risk management problems, both inside and outside mean-variance framework, and their corresponding parametric formulations in Chapter 7.

Mean-variance framework assumes normality of distribution of asset returns. Another, more attractive, possibility is to construct an approximate distribution from a set of scenarios representing possible prices for financial instruments. Effectively, the scenario-based approach uses statistical sampling to model uncertainty. In contrast to mean-variance optimization, the scenario-based approach allows for general non-normal and discrete distributions as well as for modeling of nonlinear instruments including financial derivatives. Using scenario-based approach, we develop two novel financial optimization frameworks in Chapter 6. *Portfolio replication* models are investigated in Section 6.2 and portfolio *credit risk optimization* framework is developed in Section 6.3.

1.5 Outline of the Thesis

The thesis describes the theoretical background, solution techniques and algorithms of both multiobjective optimization and parametric conic quadratic optimization as well as financial and non-financial applications of those models and algorithms. In addition, novel financial models are developed and tested. This predetermines the following organization of the thesis.

In the current Chapter 1, we make historical remarks and briefly describe continuous optimization problems and techniques. In addition, we introduce concepts of parametric and multiobjective optimization. The use of optimization in finance is briefly discussed as well. Finally, the outline of the thesis is provided.

Part I of the thesis describes theoretical, algorithmic and computational results for parametric and multiobjective optimization. Chapter 2 deals with parametric optimization algorithms for linear and convex quadratic optimization problems. Chapter 3 contains the background discussions on multiobjective optimization, and we make a link from multiobjective optimization to its parametric counterpart. Chapter 3 also describes how multiobjective problems can be solved via parametric optimization. Chapter 4 includes details on implementing parametric optimization algorithms, moreover, a number of illustrative examples of parametric optimization problems are solved there. Selected non-financial applications of multiobjective and parametric optimization are covered in Chapter 5.

Part II of the thesis contains financial and risk management models developed or extended by us as well as examples of multiobjective optimization problems in finance and their solutions by parametric optimization algorithms. Chapter 6 describes general model classes in finance that are either developed by us or closely related to our work. Examples of financial multiobjective optimization problems and their solutions are shown in Chapter 7. Chapter 7 and Appendix A also present our computational results.

We briefly discuss our preliminary results on parametric second-order cone optimization in Chapter 8. Finally, Chapter 9 contains concluding remarks and suggestions for future work.

Part I

Parametric and Multiobjective Optimization

Chapter 2

Parametric Convex Quadratic Optimization

In this chapter we present an IPM and optimal partition based technique and provide a polynomial time algorithm for conducting parametric analysis of convex quadratic optimization problems. We allow simultaneous variation in the coefficient vector of the linear term of the objective function and in the right-hand side vector of the constraints. Let λ and ϵ be the perturbation parameters. The resulting problem we intend to solve is:

$$\begin{array}{rll} \phi(\lambda,\epsilon) = & \min & f(x,\lambda) \\ & \text{s.t.} & Ax = b + \epsilon\triangle b \\ & & x \geq 0, \end{array}$$

where $f(x,\lambda)$ is linear or quadratic function of x.

The method described in this chapter partitions the parameter space into so-called invariancy regions (intervals), and at the same time provides complete description of the behavior of $\phi(\lambda,\epsilon)$ and $x^*(\lambda,\epsilon)$ on each interval. The resulting algorithm solves parametric quadratic as well as parametric linear optimization problems efficiently in polynomial time.

Parametric analysis for QO generalizes the one for LO. We describe the QO results here; specializing most of the results to LO by making $Q = 0$ is straightforward. All the results are provided for the perturbed QO problem when perturbation occurs in the right-hand-side data and the linear term of

the objective function of the (QP) problem (1.1.7), simultaneously. This model includes the single-side perturbations as its subcases.

Research on the topic was triggered when a variant of parametric QO problems was proposed by Markowitz [96]. He applied it to mean-variance portfolio analysis. The basic result for parametric quadratic optimization obtained by Markowitz is that the optimal value function $\phi(\lambda)$ (efficient frontier in financial terminology) is piecewise quadratic, and can be computed. Non-degeneracy was assumed and a variant of the simplex method was used for computations.

There are two ways to perform sensitivity analysis in the case of LO: the *optimal basis* approach and the *optimal partition* approach. Those two are based on the concepts of optimal basis produced by simplex-type methods, and the optimal partition produced by IPMs. Simplex methods were used to perform the computations in earlier studies, see, e.g., Murty [106] for a comprehensive survey. Recently research on parametric analysis was revisited from the point of view of IPMs. In the optimal basis approach the goal is to find the range of perturbations for which the given optimal basic solution remains basic optimal. In contrast, the optimal partition approach aims to determine the range of perturbations for which the optimal partition remains invariant. In the non-degenerate case, when there is a unique primal-dual optimal basic solution, the two approaches coincide. Difficulties arising in parametric analysis when the problem is degenerate are studied extensively in the LO literature, see, e.g., [79]. In case of degeneracy the optimal basis is not unique and multiple optimal solutions may exist. For degenerate LO problems, the availability of strictly complementary solutions produced by IPMs allows overcoming many difficulties associated with the use of optimal bases. Adler and Monteiro [1] pioneered the use of IPMs in parametric analysis for LO (see also Jansen et al. [79]).

Computational costs of the optimal partition approach overweight the ones of optimal basis, however, the *optimal basis* approach may produce misleading and inconsistent information on parametric analysis if there are multiple optimal basic solutions. Berkelaar, Roos and Terlaky [12] emphasized shortcomings of

using optimal bases in parametric LO. By an example they showed that different optimal bases computed by different LO packages give different optimality intervals.

Even though the optimal partition approach has higher computational costs, it produces correct results. Another advantage of the *optimal partition* approach is that it allows the development of a complete algorithm for solving parametric optimization problems.

Naturally, results obtained for parametric LO were extended to parametric QO. Berkelaar et al. [10] showed that the optimal partition approach can be generalized to the QO case by introducing tri-partition of variables instead of bi-partition. They performed sensitivity analysis for the cases when perturbation occurs either in the coefficient vector of the linear term of the objective value function or in the right-hand-side of the constraints. In Section 2.2 we show how the results obtained in Berkelaar, Roos and Terlaky [12] and Berkelaar et al. [10] are generalized further to accommodate simultaneous perturbation of the data, even in the presence of degeneracy.

Considering simultaneous perturbation provides a unified approach to parametric LO and QO problems that includes perturbation of the coefficients of the linear term in the objective function or the right-hand-side vector of the constraints as its subcases. We discuss the optimal partition approach for LO and QO in Section 2.1. The results in Sections 2.1 and 2.2 are solely based on Ghaffari-Hadigheh, Romanko and Terlaky [55].

2.1 Optimal Partition in LO and QO

Next two paragraphs, borrowed from [29], page 17-18, contain a discussion about the differences between the optimal partition and an optimal basis.

"In the common case of multiple solutions, an interior point method algorithm terminates at the analytic center of the optimal face rather than at a vertex; in some respects, through the concept of optimal partition, we can in-

terpret this situation as having determined the whole set of optimal solutions. In contrast, the choice of the solution vertex provided by the simplex method is arbitrary, and depends on factors like the pivoting rule.

Often having a basic solution that identifies a vertex is considered to be an exact solution. However, we should discuss what we mean by "exact" solution. In most cases we do not need the additional precision of being on a vertex solution rather than at the analytic center of the optimal face. In this sense, integer programming represents a notable exception, as the integer solutions are at the vertices of the convex hull of feasible integer points. The difference between having an optimal basis or an optimal partition has important consequences on the use of the solution for sensitivity analysis."

We first consider the optimal partition for the QO problem (1.1.7). The *optimal partition* of the index set $\{1, 2, \ldots, n\}$ is defined as

$$\begin{aligned} \mathcal{B} &= \{\, i : x_i > 0 \text{ for an optimal solution } x \in \mathcal{QP}^*\}, \\ \mathcal{N} &= \{\, i : s_i > 0 \text{ for an optimal solution } (x, y, s) \in \mathcal{QD}^*\}, \\ \mathcal{T} &= \{1, 2, \ldots, n\} \setminus (\mathcal{B} \cup \mathcal{N}), \end{aligned} \tag{2.1.1}$$

and denoted by $\pi = (\mathcal{B}, \mathcal{N}, \mathcal{T})$. In the definition (2.1.1) of the optimal partition $\mathcal{QP}^*$ and $\mathcal{QD}^*$ are the optimal solution sets defined on page 6. Berkelaar et al. [10] and Berkelaar, Roos and Terlaky [12] showed that this partition is unique. The *support set* of a vector v is defined as $\sigma(v) = \{i : v_i > 0\}$ and is used extensively in this chapter. An optimal solution (x, y, s) is called *maximally complementary* if it possesses the following properties:

$$\begin{aligned} x_i > 0 \quad &\text{if and only if} \quad i \in \mathcal{B}, \\ s_i > 0 \quad &\text{if and only if} \quad i \in \mathcal{N}. \end{aligned}$$

For any maximally complementary solution (x, y, s) the relations $\sigma(x) = \mathcal{B}$ and $\sigma(s) = \mathcal{N}$ hold. The existence of a maximally complementary solution is a direct consequence of the convexity of the optimal sets $\mathcal{QP}^*$ and $\mathcal{QD}^*$. It is known that IPMs find a maximally complementary solution in the limit [100, 66]. Interior point methods are widely used to solve QO problems in polynomial time [107]

and sufficiently accurate solutions obtained by an IPM can be used to produce maximally complementary solutions [73]. By knowing a maximality complementary optimal solution, one can easily identify the optimal partition. If for a given optimal partition $\mathcal{T} = \emptyset$ holds, then any maximally complementary solution is *strictly complementary*. The optimal partition in the LO case consists of two sets only $\pi = (\mathcal{B}, \mathcal{N})$ because a strictly complementary solution for LO problems always exists. It is worth mentioning that for any primal-dual optimal solution (x^*, y^*, s^*), the relations $\sigma(x^*) \subseteq \mathcal{B}$ and $\sigma(s^*) \subseteq \mathcal{N}$ hold. Both inclusions hold with equality if and only if the given primal-dual optimal solution is maximally (strictly) complementary.

The importance of the optimal partition for sensitivity and parametric analysis is that it is constant on the invariancy regions (invariancy intervals for the single-parameter case).

2.2 Single-Parametric Quadratic Optimization

The primal and dual perturbed problems corresponding to (QP) and (QD) (equations (1.1.7) and (1.1.8)), respectively, are:

$$(QP_\lambda) \qquad \begin{array}{rcll} \phi(\lambda) & = & \min & (c+\lambda\triangle c)^T x + \frac{1}{2}x^T Q x \\ & & \text{s.t.} & Ax = b + \lambda\triangle b \\ & & & x \geq 0, \end{array} \tag{2.2.1}$$

$$(QD_\lambda) \qquad \begin{array}{rll} \max & (b+\lambda\triangle b)^T y - \frac{1}{2}x^T Q x \\ \text{s.t.} & A^T y + s - Qx = c + \lambda\triangle c \\ & s \geq 0. \end{array} \tag{2.2.2}$$

The perturbation takes the form λh, where $h = (\triangle b^T, \triangle c^T)^T \in \mathbb{R}^{m+n}$ is a nonzero perturbing direction and $\lambda \in \mathbb{R}$ is a parameter.

Consider the problem (QP_λ) defined by (2.2.1). Its solution methodology is our primary interest in this section. Let $\mathcal{QP}_\lambda$ and $\mathcal{QD}_\lambda$ denote the feasible sets of problems (QP_λ) and (QD_λ) defined by (2.2.1) and (2.2.2), respectively. Their optimal solution sets are analogously denoted by $\mathcal{QP}^*_\lambda$ and $\mathcal{QD}^*_\lambda$. The optimal

value function of (QP_λ) and (QD_λ) is

$$\phi(\lambda) = (c+\lambda\triangle c)^T x^*(\lambda) + \frac{1}{2}x^*(\lambda)^T Q x^*(\lambda) = (b+\lambda\triangle b)^T y^*(\lambda) - \frac{1}{2}x^*(\lambda)^T Q x^*(\lambda),$$

where $x^*(\lambda) \in \mathcal{QP}^*_\lambda$ and $(x^*(\lambda), y^*(\lambda), s^*(\lambda)) \in \mathcal{QD}^*_\lambda$. Further, we define

$$\begin{array}{ll} \phi(\lambda) = +\infty & \text{if } \mathcal{QP}_\lambda = \emptyset, \\ \phi(\lambda) = -\infty & \text{if } \mathcal{QP}_\lambda \neq \emptyset \text{ and } (QP_\lambda) \text{ is unbounded.} \end{array}$$

Let us denote the domain of $\phi(\lambda)$ by

$$\Lambda = \{\lambda : \mathcal{QP}_\lambda \neq \emptyset \text{ and } \mathcal{QD}_\lambda \neq \emptyset\}.$$

Since it is assumed that (QP) and (QD) (problems (1.1.7) and (1.1.8)) have optimal solutions, it follows that $\Lambda \neq \emptyset$. We can prove the following property of Λ: $\Lambda \subseteq \mathbb{R}$ is a closed interval.

Denote by $\mathcal{O}(\pi) = \{\lambda \in \Lambda : \pi(\lambda) = \pi\}$ the set of parameter values for which the optimal partition π is constant. $\Lambda(\pi)$ denotes the set of parameter values where the perturbed primal and dual problems are feasible when restricting the optimal solutions to the same partition. Finally, $\overline{\Lambda}(\pi)$ refers to the closure of $\Lambda(\pi)$.

Let's say that for some λ we are given a maximally complementary optimal solution (x^*, y^*, s^*) of (QP_λ) and (QD_λ) with the optimal partition $\pi = \pi(\lambda) = (\mathcal{B}, \mathcal{N}, \mathcal{T})$.

Theorem 2.2.1 *Let $\lambda_2 > \lambda_1$ be such that $\pi(\lambda_1) = \pi(\lambda_2)$. Then, $\pi(\lambda)$ is constant for all $\lambda \in [\lambda_1, \lambda_2]$.*

Theorem 2.2.2 *Let $(x^{(1)}, y^{(1)}, s^{(1)})$ and $(x^{(2)}, y^{(2)}, s^{(2)})$ be maximally complementary solutions of (QP_{λ_1}), (QD_{λ_1}) and (QP_{λ_2}), (QD_{λ_2}), respectively. Then, for any $\lambda \in [\lambda_1, \lambda_2]$*

$$\begin{aligned} x(\lambda) &= \frac{\lambda_2 - \lambda}{\lambda_2 - \lambda_1}x^{(1)} + \frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1}x^{(2)}, \\ y(\lambda) &= \frac{\lambda_2 - \lambda}{\lambda_2 - \lambda_1}y^{(1)} + \frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1}y^{(2)}, \\ s(\lambda) &= \frac{\lambda_2 - \lambda}{\lambda_2 - \lambda_1}s^{(1)} + \frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1}s^{(2)} \end{aligned}$$

is a maximally complementary solution of (QP_λ) *and* (QD_λ) *if and only if* $\lambda_1, \lambda_2 \in \Lambda(\pi)$.

Theorem 2.2.2 shows how to identify a maximally complementary optimal solutions on $\Lambda(\pi)$. The following theorem allows us to compute the endpoints of the interval $\overline{\Lambda}(\pi)$ efficiently.

Theorem 2.2.3 *Let* $\lambda^* \in \Lambda$ *and let* (x^*, y^*, s^*) *be a maximally complementary solution of* (QP_{λ^*}) *and* (QD_{λ^*}) *with optimal partition* $\pi = (\mathcal{B}, \mathcal{N}, \mathcal{T})$. *Then the left and right extreme points of the closed interval* $\overline{\Lambda}(\pi) = [\lambda_\ell, \lambda_u]$ *that contains* λ^* *can be obtained by solving*

$$\begin{aligned} \lambda_\ell \;= \min_{\lambda,x,y,s} \{\, \lambda : \;& Ax - \lambda\triangle b = b,\; x_{\mathcal{B}} \geq 0,\; x_{\mathcal{N}\cup\mathcal{T}} = 0, \\ & A^T y + s - Qx - \lambda\triangle c = c,\; s_{\mathcal{N}} \geq 0,\; s_{\mathcal{B}\cup\mathcal{T}} = 0 \,\}, \end{aligned} \tag{2.2.3}$$

and

$$\begin{aligned} \lambda_u \;= \max_{\lambda,x,y,s} \{\, \lambda : \;& Ax - \lambda\triangle b = b,\; x_{\mathcal{B}} \geq 0,\; x_{\mathcal{N}\cup\mathcal{T}} = 0, \\ & A^T y + s - Qx - \lambda\triangle c = c,\; s_{\mathcal{N}} \geq 0,\; s_{\mathcal{B}\cup\mathcal{T}} = 0 \,\}. \end{aligned} \tag{2.2.4}$$

The open interval $\Lambda(\pi)$ is referred to as *invariancy interval* because the optimal partition is invariant on it. The points λ_ℓ and λ_u, that separate neighboring invariancy intervals, are called *transition points.*

Theorem 2.2.4 *Let* $\lambda_\ell < \lambda_u$ *be obtained by solving (2.2.3) and (2.2.4), respectively. The optimal value function* $\phi(\lambda)$ *is quadratic on* $\mathcal{O}(\pi) = (\lambda_\ell, \lambda_u)$. *Moreover, the optimal value function* $\phi(\lambda)$ *is continuous and piecewise quadratic on* Λ.

The concepts defined above are illustrated in Figure 2.1. We know how the optimal value function can be computed for the current invariancy interval. The question is how to proceed from the current invariancy interval with the optimal partition $\pi = (\mathcal{B}, \mathcal{N}, \mathcal{T})$ to the neighboring one with the optimal partition $\overline{\pi} = (\overline{\mathcal{B}}, \overline{\mathcal{N}}, \overline{\mathcal{T}})$. In order to do so, it is necessary to compute left and right derivatives in the transition point and solve the auxiliary self-dual CQO problem.

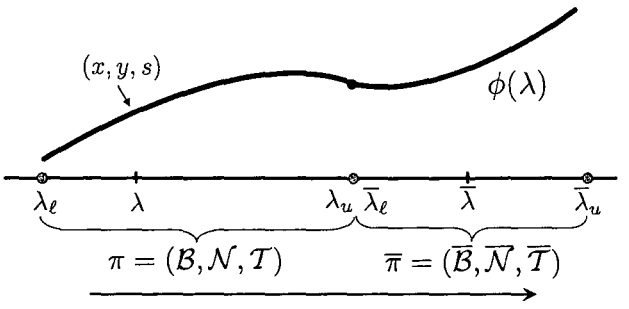


Figure 2.1: The Optimal Value Function and Invariancy Intervals of Single-Parametric QO.

Theorem 2.2.5 *For a given $\lambda \in \Lambda$, the left and right derivatives of the optimal value function $\phi(\lambda)$ at λ satisfy*

$$\phi'_-(\lambda) = \min_{x,y,s}\{\triangle b^T y : (x, y, s) \in \mathcal{QD}^*_\lambda\} + \max_x\{\triangle c^T x : x \in \mathcal{QP}^*_\lambda\}, \qquad (2.2.5)$$

$$\phi'_+(\lambda) = \max_{x,y,s}\{\triangle b^T y : (x, y, s) \in \mathcal{QD}^*_\lambda\} + \min_x\{\triangle c^T x : x \in \mathcal{QP}^*_\lambda\}. \qquad (2.2.6)$$

Next, we use basic properties of the optimal value function and its derivatives to investigate the relationship between the invariancy intervals and neighboring transition points where these derivatives may not exist. We also show how we can proceed from one invariancy interval to another one, to cover the whole interval Λ. These results allow us to develop our algorithm for solving parametric QO problems.

It is worthwhile to make first some remarks about Theorem 2.2.5. It seems that we need to solve two optimization problems to find the right or left first-order derivatives of the optimal value function at a transition point. Actually we can combine these two problems into one. We consider problem (2.2.6) only. Similar results hold for problem (2.2.5). Let (x^*, y^*, s^*) be a pair of primal-dual optimal solutions of (QP_λ) and (QD_λ) and

$$\begin{aligned}\mathcal{QPD}^*_\lambda = \{(x, y, s) : \; & Ax = b + \lambda\triangle b,\; x \geq 0,\; x^T s^* = 0,\; Qx = Qx^*, \\ & A^T y + s - Qx = c + \lambda\triangle c,\; s \geq 0,\; s^T x^* = 0\}.\end{aligned}$$

First, in the definition of the set $\mathcal{QPD}^*_\lambda$ the constraints $x \geq 0$, $x^T s^* = 0$, $Qx = Qx^*$ and $s \geq 0$, $s^T x^* = 0$ are equivalent to $x_\mathcal{B} \geq 0$, $x_{\mathcal{N}\cup\mathcal{T}} = 0$ and $s_\mathcal{N} \geq 0$, $s_{\mathcal{B}\cup\mathcal{T}} = 0$, where $(\mathcal{B}, \mathcal{N}, \mathcal{T})$ is the optimal partition at the transition point λ. The fact that $x_\mathcal{B} \geq 0$ directly follows from $x \geq 0$. On the other hand, since (x, y, s) is a primal-dual optimal solution and (x^*, y^*, s^*) is a maximally complementary optimal solution, then $\sigma(x) \subseteq \sigma(x^*)$, thus $x_{\mathcal{N}\cup\mathcal{T}} = 0$ is its immediate result. Analogous reasoning is valid for $s_{\mathcal{B}\cup\mathcal{T}} = 0$. Second, let us consider the first and the second subproblems of (2.2.6). Observe that the optimal solutions produced by each subproblem are both optimal for (QP_λ) and (QD_λ) and so the vector Qx, appearing in the constraints, is always identical for both subproblems (see, e.g., [40]). This means that we can maximize the first subproblem over $\mathcal{QPD}^*_\lambda$ and minimize the second subproblem over $\mathcal{QPD}^*_\lambda$ simultaneously. In other words, instead of solving two subproblems in (2.2.6) separately, we can solve the problem

$$\min_{x,y,s}\{\triangle c^T x - \triangle b^T y : (x, y, s) \in \mathcal{QPD}^*_\lambda\}, \tag{2.2.7}$$

that produces the same optimal solution $(\hat{x}, \hat{y}, \hat{s})$ as a solution of problem (2.2.6). Then the right derivative $\phi'_+(\lambda)$ can be computed by using the values $(\hat{x}, \hat{y}, \hat{s})$ as $\phi'_+(\lambda) = \triangle b^T \hat{y} + \triangle c^T \hat{x}$. Consequently, we refer to the optimal solutions of problems (2.2.6) and (2.2.7) interchangeably.

The next lemma shows an important property of strictly complementary solutions of (2.2.5) and (2.2.6) that is used in the proof of Theorem 2.2.7.

Lemma 2.2.6 *Let λ^* be a transition point of the optimal value function. Further, assume that the (open) invariancy interval to the right of λ^* contains $\overline{\lambda}$ with the optimal partition $\overline{\pi} = (\overline{\mathcal{B}}, \overline{\mathcal{N}}, \overline{\mathcal{T}})$. Let (x, y, s) be an optimal solution of (2.2.6) with $\lambda = \lambda^*$. Then, $\sigma(x) \subseteq \overline{\mathcal{B}}$ and $\sigma(s) \subseteq \overline{\mathcal{N}}$.*

By solving an auxiliary self-dual quadratic optimization problem defined in Theorem 2.2.7 we can obtain the optimal partition in the neighboring invariancy interval.

Theorem 2.2.7 *Let λ^* be a transition point of the optimal value function. Let (x^*, y^*, s^*) be an optimal solution of (2.2.6) for λ^*. Let us assume that the (open) invariancy interval to the right of λ^* contains $\overline{\lambda}$ with optimal partition $\overline{\pi} = (\overline{\mathcal{B}}, \overline{\mathcal{N}}, \overline{\mathcal{T}})$. Define $T = \overline{\sigma}(x^*, s^*) = \{1, 2, \ldots, n\} \setminus (\sigma(x^*) \cup \sigma(s^*))$. Consider the following self-dual quadratic problem*

$$\begin{aligned} \min_{\xi,\rho,\eta} \quad & \{-\triangle b^T\eta + \triangle c^T\xi + \xi^T Q\xi : A\xi = \triangle b,\ A^T\eta + \rho - Q\xi = \triangle c, \\ & \xi_{\sigma(s^*)} = 0,\ \rho_{\sigma(x^*)} = 0,\ \xi_{\overline{\sigma}(x^*,s^*)} \geq 0,\ \rho_{\overline{\sigma}(x^*,s^*)} \geq 0\}, \end{aligned} \tag{2.2.8}$$

and let (ξ^*, η^*, ρ^*) be a maximally complementary solution of (2.2.8). Then, $\overline{\mathcal{B}} = \sigma(x^*) \cup \sigma(\xi^*)$, $\overline{\mathcal{N}} = \sigma(s^*) \cup \sigma(\rho^*)$ and $\overline{\mathcal{T}} = \{1, \ldots, n\} \setminus (\overline{\mathcal{B}} \cup \overline{\mathcal{N}})$.

Results described in this section allow outlining Algorithm 2.1 for identifying all invariancy intervals, and computing the optimal value function and maximally complementary solutions on all those intervals. Note that the algorithm computes all these quantities to the right from the given initial value λ^*. One can easily outline an analogous algorithm for the transition points to the left from λ^*, e.g., by taking the negative of the perturbation vectors. All the subproblems used in Algorithm 2.1 can be solved in polynomial time by IPMs. Numerical results and implementation details for Algorithm 2.1 can be found in [123].

2.3 Bi-Parametric Quadratic Optimization

Bi-parametric optimization problems are the natural extension of their uni-parametric counterparts. In addition, allowing the independent variation of the objective function coefficients and the right-hand-side of the constraints cover a much wider class of models. The recent developments on the bi-parametric QO are described in [56, 57].

The *bi-parametric QO problem* is

$$\begin{aligned} (QP_{\lambda,\epsilon}) \qquad \phi(\lambda,\epsilon) \ = \ \min \quad & (c + \lambda\triangle c)^T x + \tfrac{1}{2} x^T Q x \\ \text{s.t.} \quad & Ax = b + \epsilon\triangle b \\ & x \geq 0, \end{aligned} \tag{2.3.1}$$

Algorithm 2.1: Algorithm for Enumerating All Invariancy Intervals for Single-Parametric QO Problems.

Input:
A nonzero direction of perturbation: $r = (\triangle b, \triangle c)$;
a maximally complementary solution (x^*, y^*, s^*) of (QP_λ) and (QD_λ) for $\lambda = \lambda^*$;
$\pi^0 = (\mathcal{B}^0, \mathcal{N}^0, \mathcal{T}^0)$, where $\mathcal{B}^0 = \sigma(x^*)$, $\mathcal{N}^0 = \sigma(s^*)$;
$k := 0$; $x^0 := x^*$; $y^0 := y^*$; $s^0 := s^*$;
ready:= false;

while not ready **do**
begin
solve $\lambda_k = \max_{\lambda,x,y,s}\{\lambda : Ax - \lambda\triangle b = b,\ x_{\mathcal{B}^k} \geq 0,\ x_{\mathcal{N}^k \cup \mathcal{T}^k} = 0,$
$A^T y + s - Qx - \lambda\triangle c = c,\ s_{\mathcal{N}^k} \geq 0,\ s_{\mathcal{B}^k \cup \mathcal{T}^k} = 0\}$;
if this problem is unbounded: ready:= true; **else**
let $(\lambda_k, x^k, y^k, s^k)$ be an optimal solution;
begin
Let $x^* := x^k$ and $s^* := s^k$;
solve $\min_{x,y,s}\{\triangle c^T x - \triangle b^T y : (x, y, s) \in \mathcal{QPD}^*_\lambda\}$
if this problem is unbounded: ready:= true; **else**
let (x^k, y^k, s^k) be an optimal solution;
begin
Let $x^* := x^k$ and $s^* := s^k$;
solve $\min_{\xi,\rho,\eta}\{-\triangle b^T \eta + \triangle c^T \xi + \xi^T Q \xi : A\xi = \triangle b,$
$A^T\eta + \rho - Q\xi = \triangle c,\ \xi_{\sigma(s^*)} = 0,$
$\rho_{\sigma(x^*)} = 0,\ \xi_{\bar{\sigma}(x^*,s^*)} \geq 0,\ \rho_{\bar{\sigma}(x^*,s^*)} \geq 0\}$;
$\mathcal{B}^{k+1} = \sigma(x^*) \cup \sigma(\xi^*)$, $\mathcal{N}^{k+1} = \sigma(s^*) \cup \sigma(\rho^*)$,
$\mathcal{T}^{k+1} = \{1, \ldots, n\} \setminus (\mathcal{B}^{k+1} \cup \mathcal{N}^{k+1})$;
$k := k + 1$;
end
end
end

where $A \in \mathbb{R}^{m\times n}$, $Q \in \mathbb{R}^{n\times n}$ is a symmetric semi-definite matrix, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$ are fixed data, ϵ and λ are two real parameters, $\triangle b \in \mathbb{R}^m$, $\triangle c \in \mathbb{R}^n$ are the perturbation directions and $x \in \mathbb{R}^n$ is an unknown vector. The dual of problem $(QP_{\lambda,\epsilon})$ is

$$(QD_{\lambda,\epsilon}) \qquad \begin{array}{ll} \max & (b+\epsilon\triangle b)^T y - \frac{1}{2}x^T Q x \\ \text{s.t.} & A^T y + s - Qx = c + \lambda\triangle c \\ & x, s \geq 0, \end{array} \tag{2.3.2}$$

where $y \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ are unknowns. As explained in Section 1.1.3, we may assume without loss of generality that dual variable x is equal to the primal variable x. In general, both $\triangle b$ and $\triangle c$ are non-zero vectors. Following the convention in Section 2.2, $\mathcal{QP}_{\lambda,\epsilon}$ and $\mathcal{QP}^*_{\lambda,\epsilon}$ denote the sets of primal feasible and primal optimal solutions of $(QP_{\lambda,\epsilon})$, respectively. Similar notation is used for the sets of feasible and optimal solutions of $(QD_{\lambda,\epsilon})$. Optimal solutions depend on both ϵ and λ, consequently we denote the primal-dual optimal solutions of the perturbed QO problems by $(x^*(\epsilon,\lambda), y^*(\epsilon,\lambda), s^*(\epsilon,\lambda))$.

In the context of bi-parametric QO, we denote the *optimal partition* (see definition (2.1.1)) by $\pi(\epsilon,\lambda) = (\mathcal{B}(\epsilon,\lambda), \mathcal{N}(\epsilon,\lambda), \mathcal{T}(\epsilon,\lambda))$:

$$\begin{aligned} \mathcal{B}(\epsilon,\lambda) &= \{i : x_i^*(\epsilon,\lambda) > 0 \text{ for an optimal solution } x^*(\epsilon,\lambda)\}, \\ \mathcal{N}(\epsilon,\lambda) &= \{i : s_i^* > 0 \text{ for an optimal solution } (x^*(\epsilon,\lambda), y^*(\epsilon,\lambda), s^*(\epsilon,\lambda))\}, \\ \mathcal{T}(\epsilon,\lambda) &= \{1,2,\cdots,n\} \setminus (\mathcal{B}(\epsilon,\lambda) \cup \mathcal{N}(\epsilon,\lambda)) \\ &= \{i : x_i^*(\epsilon,\lambda) = s_i^*(\epsilon,\lambda) = 0 \text{ for all primal-dual optimal solutions} \\ &\qquad\qquad (x^*(\epsilon,\lambda), y^*(\epsilon,\lambda), s^*(\epsilon,\lambda))\}. \end{aligned}$$

The *optimal value function* $\phi(\epsilon,\lambda)$ of the bi-parametric QO problem is defined as:

$$\begin{aligned} \phi(\epsilon,\lambda) &= (c+\lambda\triangle c)^T x^*(\epsilon,\lambda) + \tfrac{1}{2}x^*(\epsilon,\lambda)^T Q x^*(\epsilon,\lambda) \\ &= (b+\epsilon\triangle b)^T y^*(\epsilon,\lambda) - \tfrac{1}{2}x^*(\epsilon,\lambda)^T Q x^*(\epsilon,\lambda), \end{aligned} \tag{2.3.3}$$

where $\triangle b$ and $\triangle c$ are fixed perturbing directions. The optimal value function $\phi(\epsilon,\lambda)$ denotes the optimal value of $(QP_{\lambda,\epsilon})$ as the function of the parameters λ and ϵ.

In *optimal partition invariancy* sensitivity analysis we aim to identify the range of parameters where the optimal partition remains invariant. The cases when either $\triangle b$ or $\triangle c$ is zero has been studied in [12]. The situation when $\epsilon = \lambda$ has been investigated in [55] and is described in Section 2.2. In these cases the region of the parameter is an interval of the real line called *invariancy interval*. The just mentioned special cases are referred to as *single-* or *uni-parametric* optimal partition invariancy sensitivity analysis.

Bi-parametric optimal partition based sensitivity analysis has been studied in case of LO in [54]. Bi-parametric active set based sensitivity analysis was developed in [6, 65] for the LO and QO cases. Earlier studies (see [111, 149, 72, 110, 64] for more details) produced the ideas that are summarized in [54, 6, 65] and we refer the interested reader to consult those publications. Fundamental properties derived in this section overlap with the ones that appeared in [6] with the difference that those are based on the active-set notion and, consequently, require non-degeneracy assumption. While [6] contains theoretical derivations and numerical examples, it does not provide a complete algorithm to do systematic analysis of bi-parametric QO problems. In the remainder of this paragraph, we summarize the most recent results from [54] in a nutshell. The crucial difference of the bi-parametric LO case from the QO case is that in the LO case the invariancy regions are relatively open rectangles while in the QO case those are open convex polyhedrons. In bi-parametric LO the invariancy regions generate a mesh-like area in $\mathbb{R}^2$ that simplifies enumeration of the regions. This is not the case for bi-parametric QO problems that results in a more complicated computational algorithm.

In this thesis, we consider the bi-parametric optimal partition invariancy sensitivity analysis for QO in the general case, when both $\triangle b$ and $\triangle c$ are nonzero vectors and parameters ϵ and λ change independently. Let $\pi = (\mathcal{B}, \mathcal{N}, \mathcal{T})$ denote the optimal partition for $\epsilon = 0$ and $\lambda = 0$. We are interested in finding all the regions on the "$\epsilon - \lambda$" plane where the optimal partition is invariant, i.e., $\pi(\epsilon, \lambda) = (\mathcal{B}, \mathcal{N}, \mathcal{T})$. We call each of these regions *invariancy region* and denote

it by $\mathcal{IR}(\triangle b, \triangle c)$. It is obvious that one of these regions includes the origin $(0,0)$, and thus, their union is a nonempty set.

Fundamental Properties

In this section, we prove some fundamental properties of the invariancy region and describe the behavior of the optimal value function in this region. First we prove that this region is a convex set. Analogous result can be found in Theorems 5.4.3 and 5.4.4 in [6] and Theorem 17 in [64].

Lemma 2.3.1 *The set $\mathcal{IR}(\triangle b, \triangle c)$ is a convex set and its closure is a polyhedron.*

Proof. To provide the illustration and insight we give the proof in one direction and refer to [12] for another direction. Let (ϵ_1, λ_1) and (ϵ_2, λ_2) are two arbitrary pairs in $\mathcal{IR}(\triangle b, \triangle c)$. Let $(x^{(1)}, y^{(1)}, s^{(1)})$ and $(x^{(2)}, y^{(2)}, s^{(2)})$ are maximally (strictly) complementary optimal solutions of problems $(QP_{\lambda,\epsilon})$ and $(QD_{\lambda,\epsilon})$ at these points. Let (ϵ, λ) be an arbitrary point on the line segment between the two points (ϵ_1, λ_1) and (ϵ_2, λ_2). There is a $\theta \in (0,1)$ such that:

$$\epsilon = \epsilon_1 + \theta \triangle \epsilon, \tag{2.3.4}$$

$$\lambda = \lambda_1 + \theta \triangle \lambda, \tag{2.3.5}$$

where $\triangle \epsilon = \epsilon_2 - \epsilon_1$ and $\triangle \lambda = \lambda_2 - \lambda_1$. We define

$$x(\epsilon, \lambda) = \theta x^{(1)} + (1-\theta) x^{(2)}, \tag{2.3.6}$$

$$y(\epsilon, \lambda) = \theta y^{(1)} + (1-\theta) y^{(2)}, \tag{2.3.7}$$

$$s(\epsilon, \lambda) = \theta s^{(1)} + (1-\theta) s^{(2)}. \tag{2.3.8}$$

It is easy to verify that $x(\epsilon, \lambda)$ is a primal feasible solution and $(x(\epsilon, \lambda), y(\epsilon, \lambda), s(\epsilon, \lambda))$ is a dual feasible solution. On the other hand, $\sigma(x(\epsilon, \lambda)) = \sigma(x^{(1)}) \cup \sigma(x^{(2)}) = \mathcal{B}$ and $\sigma(s(\epsilon, \lambda)) = \sigma(s^{(1)}) \cup \sigma(s^{(2)}) = \mathcal{N}$, that proves the optimality of these solutions for problems $(QP_{\lambda,\epsilon})$ and $(QD_{\lambda,\epsilon})$. This implies that

$$\mathcal{B} \subseteq \mathcal{B}_{\epsilon,\lambda}, \ \mathcal{N} \subseteq \mathcal{N}_{\epsilon,\lambda} \text{ and } \mathcal{T} \supseteq \mathcal{T}_{\epsilon,\lambda}.$$

To show that equality holds, we assume to the contrary that $\mathcal{T} \supset \mathcal{T}_{\epsilon,\lambda}$ and use a contradiction argument from the proof of Theorem 6.40 in [12]. It establishes that the optimal partition is $\pi = (\mathcal{B}, \mathcal{N}, \mathcal{T})$ at (ϵ, λ). That completes the proof of the first statement. Having the optimal partition given, the optimality conditions reduce to a linear inequality system when one fixes all x_i variables zero in the $\mathcal{N} \cup \mathcal{T}$ part and the s_i variables zero at the $\mathcal{B} \cup \mathcal{T}$ part, giving a polyhedron [120]. □

The boundaries between the invariancy regions are line (half-line) segments. The line segment between two adjacent invariancy regions is referred to as *transition line segment* and the intersection of two transition lines are called *transition points*. Transition points (singleton invariancy regions) and transition lines are called *trivial* invariancy regions. An invariancy region that is neither a singleton nor a transition line segment is referred to as *non-trivial* invariancy region.

The optimal value function $\phi(\epsilon, \lambda)$ is continuous and piecewise-quadratic. While these results are proven in Theorems 5.5.1 and 5.5.2 in [6], the expression for quadratic function given in the proof below is not derived in [6].

Theorem 2.3.2 *The optimal value function is a bivariate quadratic function on any invariancy region $\mathcal{IR}(\triangle b, \triangle c)$.*

Proof. If the invariancy region is a non-singleton trivial region, then the optimal value function is a univariate quadratic function by Theorem 4.5 in [55]. Let the invariancy region be a non-trivial region. Further, let (ϵ_1, λ_1), (ϵ_2, λ_2) and (ϵ_3, λ_3) are three points in general position (not on a line) on the "$\epsilon - \lambda$" plane. We are allowed to make assumption about the general position of points since the invariancy region is not a trivial region. Let $(x^*(\epsilon_1, \lambda_1), y^*(\epsilon_1, \lambda_1), s^*(\epsilon_1, \lambda_1))$, $(x^*(\epsilon_2, \lambda_2), y^*(\epsilon_2, \lambda_2), s^*(\epsilon_2, \lambda_2))$ and $(x^*(\epsilon_3, \lambda_3), y^*(\epsilon_3, \lambda_3), s^*(\epsilon_3, \lambda_3))$ be primal-dual optimal solutions at these three points, respectively. Moreover, let (ϵ, λ) be an arbitrary point in the interior of the triangle formed by these three points. Therefore, there are $\theta_1, \theta_2 \in (0, 1)$ with $0 < \theta_1 + \theta_2 < 1$ such that

$$\epsilon = \epsilon_3 - \theta_1 \triangle \epsilon_1 - \theta_2 \triangle \epsilon_2, \tag{2.3.9}$$

$$\lambda = \lambda_3 - \theta_1 \triangle \lambda_1 - \theta_2 \triangle \lambda_2, \tag{2.3.10}$$

where $\triangle\epsilon_1 = \epsilon_3 - \epsilon_1$, $\triangle\epsilon_2 = \epsilon_3 - \epsilon_2$, $\triangle\lambda_1 = \lambda_3 - \lambda_1$ and $\triangle\lambda_2 = \lambda_3 - \lambda_2$. Let us define

$$\begin{aligned} x^*(\epsilon,\lambda) &= x^{(3)} - \theta_1\triangle x^{(1)} - \theta_2\triangle x^{(2)}, \\ y^*(\epsilon,\lambda) &= y^{(3)} - \theta_1\triangle y^{(1)} - \theta_2\triangle y^{(2)}, \\ s^*(\epsilon,\lambda) &= s^{(3)} - \theta_1\triangle s^{(1)} - \theta_2\triangle s^{(2)}, \end{aligned} \tag{2.3.11}$$

where $\triangle x^{(1)} = x^{(3)} - x^{(1)}$, $\triangle x^{(2)} = x^{(3)} - x^{(2)}$, $\triangle y^{(1)} = y^{(3)} - y^{(1)}$, $\triangle y^{(2)} = y^{(3)} - y^{(2)}$, $\triangle s^{(1)} = s^{(3)} - s^{(1)}$ and $\triangle s^{(2)} = s^{(3)} - s^{(2)}$. It is easy to verify that $(x^*(\epsilon,\lambda), y^*(\epsilon,\lambda), s^*(\epsilon,\lambda))$ is a primal-dual optimal solution of problems $(QP_{\lambda,\epsilon})$ and $(QD_{\lambda,\epsilon})$. Substituting (2.3.10) and (2.3.11) in (2.3.3) gives

$$\begin{aligned} \phi(\epsilon,\lambda) &= (b+\epsilon\triangle b)^T y^*(\epsilon,\lambda) - \frac{1}{2}x^*(\epsilon,\lambda)^T Q x^*(\epsilon,\lambda) \\ &= a_0 + a_1\theta_1 + a_2\theta_2 + a_3\theta_1\theta_2 + a_4\theta_1^2 + a_5\theta_2^2, \end{aligned} \tag{2.3.12}$$

where

$$\begin{aligned} a_0 &= b^T y^{(3)} + \epsilon_3\triangle b^T y^{(3)} - \tfrac{1}{2}(x^{(3)})^T Q x^{(3)}, \\ a_1 &= -b^T\triangle y^{(1)} - \epsilon_3\triangle b^T y^{(1)} - \triangle\epsilon_1\triangle b^T y^{(3)} + (x^{(3)})^T Q\triangle x^{(1)}, \\ a_2 &= -b^T\triangle y^{(2)} - \epsilon_3\triangle b^T y^{(2)} - \triangle\epsilon_2\triangle b^T y^{(3)} + (x^{(3)})^T Q\triangle x^{(2)}, \\ a_3 &= \triangle\epsilon_1\triangle b^T\triangle y^{(2)} + \triangle\epsilon_2\triangle b^T\triangle y^{(1)} - (\triangle x^{(2)})^T Q\triangle x^{(2)}, \\ a_4 &= \triangle\epsilon_1\triangle b^T\triangle y^{(1)} - \tfrac{1}{2}(\triangle x^{(1)})^T Q\triangle x^{(1)}, \\ a_5 &= \triangle\epsilon_2\triangle b^T\triangle y^{(2)} - \tfrac{1}{2}(\triangle x^{(2)})^T Q\triangle x^{(2)}. \end{aligned} \tag{2.3.13}$$

On the other hand, solving equations (2.3.10) and (2.3.9) for θ_1 and θ_2 gives

$$\theta_1 = \alpha_1 + \beta_1\epsilon + \gamma_1\lambda, \tag{2.3.14}$$

$$\theta_2 = \alpha_2 + \beta_2\epsilon + \gamma_2\lambda, \tag{2.3.15}$$

where

$$\begin{aligned} \alpha_1 &= \tfrac{\epsilon_3\triangle\lambda_2 - \lambda_3\triangle\epsilon_2}{\triangle\epsilon_1\triangle\lambda_2 - \triangle\epsilon_2\triangle\lambda_1}, & \alpha_2 &= \tfrac{\lambda_3(\triangle\epsilon_1+\triangle\epsilon_2) - \epsilon_3(\triangle\lambda_1+\triangle\lambda_2)}{\triangle\epsilon_1\triangle\lambda_2 - \triangle\epsilon_2\triangle\lambda_1}, \\ \beta_1 &= \tfrac{-\triangle\lambda_2}{\triangle\epsilon_1\triangle\lambda_2 - \triangle\epsilon_2\triangle\lambda_1}, & \beta_2 &= \tfrac{(\triangle\lambda_2+\triangle\lambda_1)}{\triangle\epsilon_1\triangle\lambda_2 - \triangle\epsilon_2\triangle\lambda_1}, \\ \gamma_1 &= \tfrac{\triangle\epsilon_2}{\triangle\epsilon_1\triangle\lambda_2 - \triangle\epsilon_2\triangle\lambda_1}, & \gamma_2 &= \tfrac{-(\triangle\epsilon_2+\triangle\epsilon_1)}{\triangle\epsilon_1\triangle\lambda_2 - \triangle\epsilon_2\triangle\lambda_1}. \end{aligned} \tag{2.3.16}$$

Substituting (2.3.13)-(2.3.16) in (2.3.12) leads to the following representation of the optimal value function:

$$\phi(\epsilon, \lambda) = b_0 + b_1\,\epsilon + b_2\,\lambda + b_3\,\epsilon\,\lambda + b_4\epsilon^2 + b_5\,\lambda^2, \tag{2.3.17}$$

where

$$\begin{aligned}
b_0 &= a_0 + a_1\alpha_1 + a_2\alpha_2 + a_3\alpha_1\alpha_2 + a_4\alpha_1^2 + a_5\alpha_2^2,\\
b_1 &= a_1\beta_1 + a_2\beta_2 + a_3(\alpha_1\beta_2 + \alpha_2\beta_1) + 2a_4\alpha_1\beta_1 + 2a_5\alpha_2\beta_2,\\
b_2 &= a_1\gamma_1 + a_2\gamma_2 + a_3(\alpha_1\gamma_2 + \alpha_2\gamma_1) + 2a_4\alpha_1\gamma_1 + 2a_5\alpha_2\gamma_2,\\
b_3 &= a_3(\gamma_1\beta_2 + \gamma_2\beta_1) + 2a_4\gamma_1\beta_1 + 2a_5\gamma_2\beta_2,\\
b_4 &= a_3\,\beta_1\,\beta_2 + a_4\,{\beta_1}^2 + a_5\,{\beta_2}^2,\\
b_5 &= a_3\,\gamma_1\,\gamma_2 + a_4\,{\gamma_1}^2 + a_5\,{\gamma_2}^2.
\end{aligned}$$

Clearly (2.3.17) is a quadratic function of ϵ and λ. Because (ϵ_1, λ_1), (ϵ_2, λ_2) and (ϵ_3, λ_3) are three arbitrary points in the non-trivial invariancy region, the claim of the theorem follows directly from (2.3.17). The proof is complete. □

Corollary 2.3.3 *The boundary of a non-trivial invariancy region consists of a finite number of line segments, and on each such line segment the optimal value function is a univariate quadratic function.*

Proof. We know that the optimal value function on the open set representing the non-trivial invariancy region is quadratic and represented by (2.3.17). By Lemma 2.3.1 the closure of the non-trivial invariancy region is a convex polyhedron. Thus, the non-trivial invariancy region boundary consists of a finite number of half-lines and/or line segments. On each line segment an optimal partition is defined by construction. The bi-parametric QO problem can be considered as a uni-parametric problem on each line segment by computing appropriate linear relations between two parameters. Using continuity of the optimal value function we get that in the limit for the border line segment, we also obtain a single quadratic function. So, the optimal value function on each border line segment of the non-trivial invariancy region boundary is a single univariate quadratic function. □

An important tool in identifying invariancy regions $\mathcal{IR}(\triangle b, \triangle c)$ will be detecting invariancy line segments on a line. When, e.g., $\epsilon = \lambda$ in the simplest case, the range of parameter variation is an interval of the real line. In this case one can identify the range of parameters via solving the following two auxiliary LO problems [55]:

$$\begin{aligned} \lambda_\ell \; = \min_{\lambda,x,y,s} \{\, \lambda : \;\; & Ax - \lambda\triangle b = b,\; x_{\mathcal{B}} \geq 0,\; x_{\mathcal{N}\cup\mathcal{T}} = 0, \\ & A^T y + s - Qx - \lambda\triangle c = c,\; s_{\mathcal{N}} \geq 0,\; s_{\mathcal{B}\cup\mathcal{T}} = 0 \,\}, \end{aligned} \tag{2.3.18}$$

and

$$\begin{aligned} \lambda_u \; = \max_{\lambda,x,y,s} \{\, \lambda : \;\; & Ax - \lambda\triangle b = b,\; x_{\mathcal{B}} \geq 0,\; x_{\mathcal{N}\cup\mathcal{T}} = 0, \\ & A^T y + s - Qx - \lambda\triangle c = c,\; s_{\mathcal{N}} \geq 0,\; s_{\mathcal{B}\cup\mathcal{T}} = 0 \,\}. \end{aligned} \tag{2.3.19}$$

where $\pi = (\mathcal{B}, \mathcal{N}, \mathcal{T})$ is the optimal partition for $\epsilon = \lambda = 0$.

Remark 2.3.4 *These auxiliary problems are simpler when either $\triangle b$ or $\triangle c$ is a zero vector. One may find the details in [12].*

Remark 2.3.5 *By analyzing the solutions of (2.3.18) and (2.3.19) we can make some conclusions about the invariancy regions in the "$\epsilon - \lambda$" plane. If solving problems (2.3.18) and (2.3.19) lead to the singleton $\{0\}$, then the invariancy region $\mathcal{IR}(\triangle b, \triangle c)$ is a one dimensional set or the singleton $\{0\}$. In both of the cases, the $\mathcal{IR}(\triangle b, \triangle c)$ is a trivial region. If $\lambda_\ell < \lambda_u$ in (2.3.18)-(2.3.19), then the region $\mathcal{IR}(\triangle b, \triangle c)$ is a one dimensional set or it is two dimensional, i.e. a non-trivial region.*

The invariancy region that contains the origin $(\epsilon, \lambda) = (0, 0)$ is referred to as the actual invariancy region.

Detecting the Boundary of an Invariancy Region

In this section, we describe the tools to identify a non-trivial invariancy region. Recall that for $\epsilon = \lambda$, the bi-parametric QO problem reduces to uni-parametric

QO problem. This trivial observation suggests choosing a method to convert the bi-parametric QO problem into uni-parametric QO problems. We start with finding some points on the boundary of the invariancy region. To accomplish this, we select the lines passing through the origin as

$$\lambda = t\epsilon. \tag{2.3.20}$$

For now, we assume that the slope t is positive. Substituting (2.3.20) into the problem $(QP_{\lambda,\epsilon})$ converts it to the following uni-parametric QO problem:

$$\min\left\{(c+\epsilon\overline{\triangle c})^T x + \frac{1}{2}x^T Q x \mid Ax = b + \epsilon\triangle b,\ x \geq 0\right\}, \tag{2.3.21}$$

where $\overline{\triangle c} = t\triangle c$. This way we can solve two associated auxiliary LO problems (2.3.18) and (2.3.19) to identify the range of variation for parameter ϵ when equation (2.3.20) holds. These two auxiliary LO problems are:

$$\begin{aligned}\lambda_\ell \ = \min_{\lambda,x,y,s}\{\,\lambda :\ & Ax - \lambda\triangle b = b,\ x_{\mathcal{B}} \geq 0,\ x_{\mathcal{N}\cup\mathcal{T}} = 0, \\ & A^T y + s - Qx - \lambda\overline{\triangle c} = c,\ s_{\mathcal{N}} \geq 0,\ s_{\mathcal{B}\cup\mathcal{T}} = 0\,\},\end{aligned} \tag{2.3.22}$$

and

$$\begin{aligned}\lambda_u \ = \max_{\lambda,x,y,s}\{\,\lambda :\ & Ax - \lambda\triangle b = b,\ x_{\mathcal{B}} \geq 0,\ x_{\mathcal{N}\cup\mathcal{T}} = 0, \\ & A^T y + s - Qx - \lambda\overline{\triangle c} = c,\ s_{\mathcal{N}} \geq 0,\ s_{\mathcal{B}\cup\mathcal{T}} = 0\,\},\end{aligned} \tag{2.3.23}$$

where $\pi = (\mathcal{B},\mathcal{N},\mathcal{T})$ is the optimal partition for $\epsilon = \lambda = 0$.

Let us consider the case when the problem (2.3.22) is bounded for two given distinct nonzero values t_1 and t_2 (if the problem is unbounded, we know that the invariancy region is unbounded) and their objective values are $\lambda(t_1)$ and $\lambda(t_2)$, respectively. Thus, two points $(\epsilon(t_1),\lambda(t_1))$ and $(\epsilon(t_2),\lambda(t_1))$ of the invariancy region boundary are known.

Let $\pi^1 = (\mathcal{B}(\lambda(t_1)), \mathcal{N}(\lambda(t_1)), \mathcal{T}(\lambda(t_1)))$ and $\pi^2 = (\mathcal{B}(\lambda(t_2)), \mathcal{N}(\lambda(t_2)), \mathcal{T}(\lambda(t_2)))$ be optimal partitions at $(\epsilon(t_1),\lambda(t_1))$ and $(\epsilon(t_2),\lambda(t_2))$, respectively. There are two possibilities for these optimal partitions.

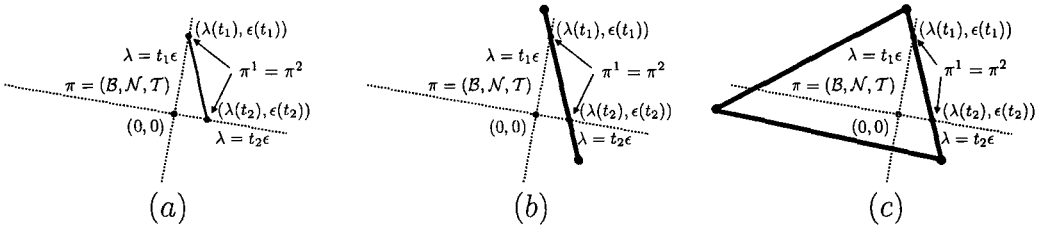


Figure 2.2: Illustration of Case 1 of the Bi-Parametric QO Algorithm.

□ **Case 1:** $\pi^1 = \pi^2$. The following lemma states that in this case we are able to identify a part of the boundary of the invariancy region that is a line on the "$\epsilon - \lambda$" plane.

Lemma 2.3.6 *Let (ϵ_1, λ_1) and (ϵ_2, λ_2) be two distinct points on the "$\epsilon - \lambda$" plane, where the optimal partitions at these points are identical. Then, the optimal partition for any point on the line segment connecting these two points is invariant, and equal to the optimal partition at the two given points.*

Proof. Let $\alpha\epsilon + \beta\lambda = 1$ be the line passing through the two points (ϵ_1, λ_1) and (ϵ_2, λ_2). We take both α and β to be non-zero, otherwise we are back to the uni-parametric case described in [55]. So, we have:

$$\lambda = \frac{1}{\beta} - \frac{\alpha}{\beta}\epsilon. \tag{2.3.24}$$

Substitution of (2.3.24) into the QO problem $(QP_{\lambda,\epsilon})$ reduces it to a uni-parametric QO problem as follows:

$$\min = \left\{ (\overline{c} + \epsilon\overline{\triangle c})^T x + \frac{1}{2} x^T Q x \mid Ax = b + \epsilon \triangle b,\ x \geq 0 \right\}, \tag{2.3.25}$$

where $\overline{c} = c + \frac{1}{\beta}\triangle c$ and $\overline{\triangle c} = \frac{\alpha}{\beta}\triangle c$. It is proven that the range of parameter variation in this case is an interval of the real line and for any ϵ in this range the optimal partition is invariant [55]. The proof is complete. □

Using Lemma 2.3.6 we can conclude that the boundary of the invariancy region contains a segment of the line (2.3.24). To identify the end points of this line segment, we need to find the invariancy interval for problem (2.3.25). The

following theorem provides two auxiliary QO problems to achieve it. The proof is straightforward and is omitted.

Theorem 2.3.7 *Let (ϵ_1, λ_1) and (ϵ_2, λ_2) be two distinct points on the "$\epsilon - \lambda$" plane. Let $\overline{\pi} = (\overline{\mathcal{B}}, \overline{\mathcal{N}}, \overline{\mathcal{T}})$ denote the optimal partition at these two points. Moreover, let $\alpha\epsilon + \beta\lambda = 1$ denote the line passing through these two points. By solving the following two auxiliary LO problems:*

$$\begin{aligned} \epsilon_\ell(\alpha,\beta) \;=\; \min_{\epsilon,x,y,s}\{\,\epsilon:\;& Ax - \epsilon\triangle b = b,\; x_{\overline{\mathcal{B}}} \geq 0,\; x_{\overline{\mathcal{N}}\cup\overline{\mathcal{T}}} = 0, \\ & A^T y + s - Qx - \epsilon\overline{\triangle c} = \overline{c},\; s_{\overline{\mathcal{N}}} \geq 0,\; s_{\overline{\mathcal{B}}\cup\overline{\mathcal{T}}} = 0\,\}, \end{aligned} \tag{2.3.26}$$

and

$$\begin{aligned} \epsilon_u(\alpha,\beta) \;=\; \max_{\epsilon,x,y,s}\{\,\epsilon:\;& Ax - \epsilon\triangle b = b,\; x_{\overline{\mathcal{B}}} \geq 0,\; x_{\overline{\mathcal{N}}\cup\overline{\mathcal{T}}} = 0, \\ & A^T y + s - Qx - \epsilon\overline{\triangle c} = \overline{c},\; s_{\overline{\mathcal{N}}} \geq 0,\; s_{\overline{\mathcal{B}}\cup\overline{\mathcal{T}}} = 0\,\}. \end{aligned} \tag{2.3.27}$$

where $\overline{c} = c + \frac{1}{\beta}\triangle c$ and $\overline{\triangle c} = \frac{\alpha}{\beta}\triangle c$, one can identify the two vertices of the invariancy region as $(\epsilon_\ell(\alpha,\beta), \lambda_\ell(\alpha,\beta))$ and $(\epsilon_u(\alpha,\beta), \lambda_u(\alpha,\beta))$, where $\lambda_\ell(\alpha,\beta) = \frac{1}{\beta} - \frac{\alpha}{\beta}\epsilon_\ell(\alpha,\beta)$ and $\lambda_u(\alpha,\beta) = \frac{1}{\beta} - \frac{\alpha}{\beta}\epsilon_u(\alpha,\beta)$.

Remark 2.3.8 *Observe that one of these auxiliary LO problems can be unbounded. In this case, the actual invariancy region is unbounded. If both problems (2.3.26) and (2.3.27) are unbounded than there is only one invariancy region on the "$\epsilon - \lambda$" plane.*

Case 1 is illustrated at Figure 2.2. Figure 2.2(a) shows two points with identical optimal partition and Figure 2.2(b) depicts the corresponding identified transition line segment. Figure 2.2(c) illustrates the non-trivial invariancy region with its boundary.

□ **Case 2:** $\pi^1 \neq \pi^2$. Let $(\overline{\epsilon}, \overline{\lambda})$ be an arbitrary point on the line segment between the two points (ϵ_1, λ_1) and (ϵ_2, λ_2). Moreover, let $\overline{\pi} = (\overline{\mathcal{B}}, \overline{\mathcal{N}}, \overline{\mathcal{T}})$ denote the associated optimal partition at $(\overline{\epsilon}, \overline{\lambda})$. We distinguish three cases for these optimal partitions.

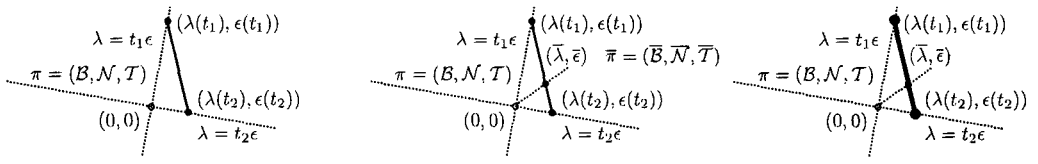


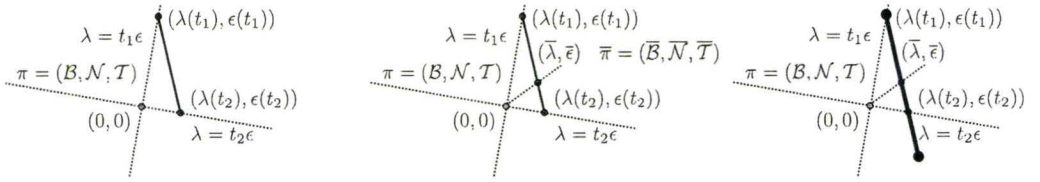
Figure 2.3: Illustration of Case 2.1 of the Bi-Parametric QO Algorithm.

Subcase 2.1: $\overline{\pi} \neq \pi$, $\overline{\pi} \neq \pi^1$ and $\overline{\pi} \neq \pi^2$. In this case, all three points (ϵ_1, λ_1), (ϵ_2, λ_2) and $(\overline{\epsilon}, \overline{\lambda})$ are on the boundary of the invariancy region. Moreover, the points (ϵ_1, λ_1) and (ϵ_2, λ_2) are the transition points on the boundary of the invariancy region. Figure 2.3 illustrates this case. The statement follows directly from Corollary 2.3.3.

Subcase 2.2: $\overline{\pi} \neq \pi$ and either $\overline{\pi} = \pi^1$ or $\overline{\pi} = \pi^2$ holds. Without loss of generality, let $\overline{\pi} \neq \pi^1$, but $\overline{\pi} = \pi^2$. In this case, both points (ϵ_2, λ_2) and $(\overline{\epsilon}, \overline{\lambda})$ are on a single boundary line of the invariancy region and consequently, Theorem 2.3.7 can be used to identify two vertices of the invariancy region. We claim that in this case, (ϵ_1, λ_1) is one of the endpoints on this part of the invariancy region. Because, if (ϵ_1, λ_1) is not one of the endpoints of this line segment, then the endpoint should be somewhere between (ϵ_1, λ_1) and $(\overline{\epsilon}, \overline{\lambda})$. It means that (ϵ_1, λ_1) is not on the line segment between (ϵ_2, λ_2) and $(\overline{\epsilon}, \overline{\lambda})$. It is obvious that (ϵ_1, λ_1) and $(\overline{\epsilon}, \overline{\lambda})$ are not in the neighboring line segment, because two adjacent line segments could not have the same slope. Thus, these two points lay on different line segments. This situation contradicts the convexity of the invariancy region. This subcase is illustrated on Figure 2.4.

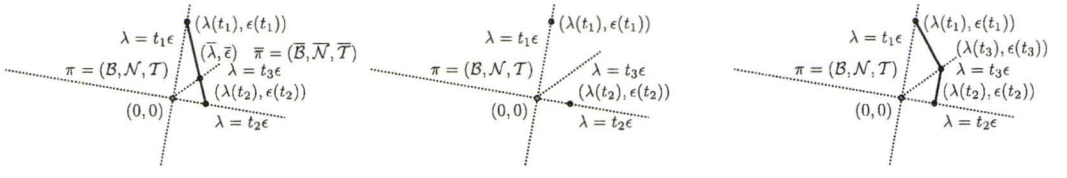
Subcase 2.3: $\overline{\pi} = \pi$. It this case, it is immediately understood that the point $(\overline{\epsilon}, \overline{\lambda})$ belongs to the invariancy region $\mathcal{IR}(\triangle b, \triangle c)$. It this case, we solve problems (2.3.22) and (2.3.23) for $\overline{\triangle c} = t_3 \triangle c$, where $t_3 = \frac{\overline{\lambda}}{\overline{\epsilon}}$.

First we prove that problem (2.3.22) for $\overline{\triangle c} = t_3 \triangle c$ is not unbounded. To



(a) (b) (c)

Figure 2.4: Illustration of Case 2.2 of the Bi-Parametric QO Algorithm.



(a) (b) (c)

Figure 2.5: Illustration of Case 2.3 of the Bi-Parametric QO Algorithm.

the contrary, let (2.3.22) is unbounded. One can consider that $t_1 = \kappa t_3$, where $\kappa > 0$. In this case, it leads to the conclusion that problem (2.3.26) is unbounded which contradicts the assumption. Thus, let ϵ_3 be the optimal objective function value of problem (2.3.22). Consequently λ_3 is known. Let $\pi^3 = (\mathcal{B}(\lambda(t_3)), \mathcal{N}(\lambda(t_3)), \mathcal{T}(\lambda(t_3)))$ be the optimal partition at (ϵ_3, λ_3). Henceforth, the situation between optimal partitions π^1 and π^3 and optimal partitions π^2 and π^3 would fall in one of the two cases: Case 1 or 2. Now, we consider pairs of points (ϵ_1, λ_1) - (ϵ_3, λ_3) and (ϵ_3, λ_3) - (ϵ_2, λ_2) again to determine if Case 1 or 2 applies to each pair. We repeat this procedure until the invariancy region boundary between the points (ϵ_1, λ_1) and (ϵ_2, λ_2) is completely traced. Figure 2.5 illustrates the procedure. In this way, we can continue to identify all the transition points (and transition lines) of the invariancy region. Since the number of optimal partitions is finite, and all auxiliary LO problems can be solved in polynomial time, thus identifying the borders of the actual invariancy region is done in polynomial time in the number of optimal partitions.

Now, we can summarize the procedure of identifying all transition points (vertices) and transition lines (edges) in an invariancy region. Lets assume that we know an initial inner point of the invariancy region and one of the edges (Figure 2.6(a) and (b) shows how to find an inner point of the region). We are going to "shoot" by solving problem (2.3.26) or (2.3.27) counter-clockwise from the initial point to identify each edge (see Figure 2.6(c-f)). As we already know one of the edges, we exclude all the angles α_{exp} between the initial point and the two vertices v_1 and v_2 of the known edge from the candidate angles to shoot. So, we shoot in the angle $v_0 - v_2$ plus in the small angles β and 2β and identify the optimal partition in the two points we get. Here we use Case 1 or Case 2 described above to find the invariancy region boundary between the vertex v_2 and the point we get when shooting in the angle 2β. If the optimal partition is the same for the points in the directions β and 2β, we compute the vertices of this new edge e_2 and verify if one of those correspond to a vertex of the previously known edge e_1. If it is not the case, then bisection is used to identify the missing edges between e_1 and e_2. We continue in this manner until all edges of the invariancy region are identified.

Transition from an Invariancy Region to the Adjacent Invariancy Regions

The initialization step of the algorithm is to get an initial inner point of an invariancy region. This step can be combined with identifying the parameters' feasibility bounds in the "$\epsilon - \lambda$" space. Alternatively, parameters' infeasibility regions can be identified during enumeration of invariancy regions. We describe only the first approach in more detail as the second one is its straightforward modification.

The first step of the algorithm is to determine the bounding box for the values of ϵ. Due to the fact that ϵ is the parameter appearing in the constraints, the problem $(QP_{\lambda,\epsilon})$ may become infeasible for large or small ϵ values. Determining the bounding box is done as in many computational geometry algo-

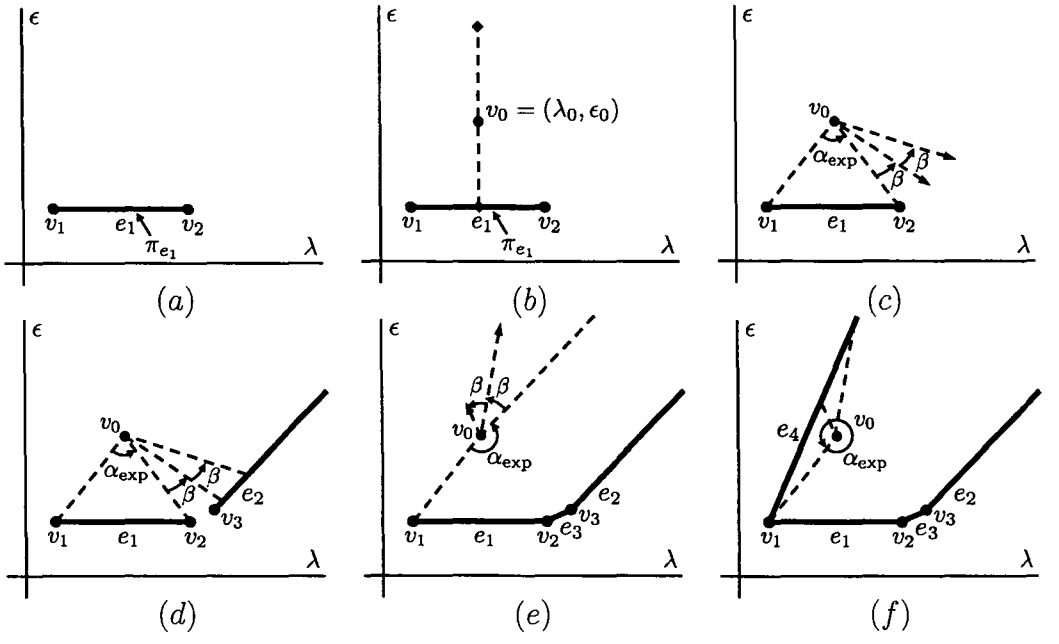


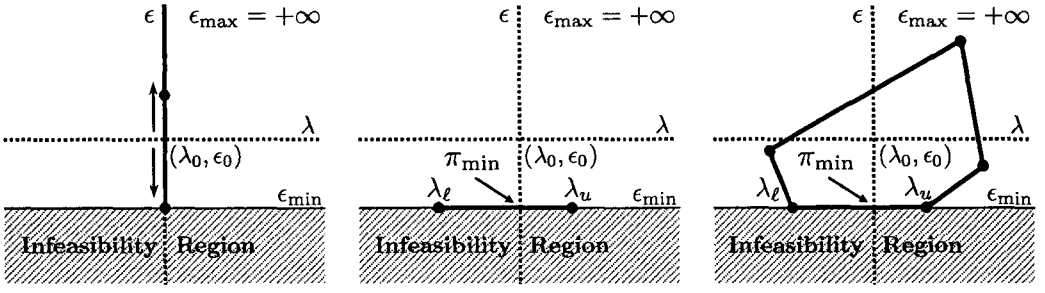
Figure 2.6: Invariancy Region Exploration Algorithm for Bi-Parametric QO.

rithms [36, 114]. To find the range of ϵ where the parametric problem $(QP_{\lambda,\epsilon})$ is feasible, we solve the following uni-parametric problem with the parameter ϵ starting from the initial point (λ_0, ϵ_0):

$$\min \left\{ (c + \triangle c\lambda_0)^T x + \frac{1}{2} x^T Q x \mid Ax = b + \epsilon \triangle b,\ x \geq 0 \right\}. \tag{2.3.28}$$

Solving problem (2.3.28) with Algorithm 2.1 gives all transition points and invariancy intervals on the line (λ_0, ϵ), and, consequently, the values of $\epsilon_{\min}$ and $\epsilon_{\max}$ that are the lower and the upper feasibility bounds for the bi-parametric problem $(QP_{\lambda,\epsilon})$, see Figure 2.7(a). Observe that we may have either $\epsilon_{\min} = -\infty$ or $\epsilon_{\max} = +\infty$, or both.

After identifying the feasibility bounds in the "$\epsilon - \lambda$" plane, we choose $\epsilon_{\min} \neq -\infty$ or $\epsilon_{\max} \neq \infty$. Let $\epsilon = \epsilon_{\min}$ and the optimal partition at the point $(\lambda_0, \epsilon_{\min})$ is $\pi_{\min} = (\mathcal{B}_{\min}, \mathcal{N}_{\min}, \mathcal{T}_{\min})$. Then we can solve problems (2.3.18) and (2.3.19) with the optimal partition $\pi = \pi_{\min}$ and $\lambda \triangle c$ replaced by $\epsilon_{\min} \triangle c$ to identify the edge on the line $\epsilon = \epsilon_{\min}$, see Figure 2.7(b). If the point $(\lambda_0, \epsilon_{\min})$ is



(*a*) (*b*) (*c*)

Figure 2.7: The Initialization of the Bi-Parametric QO Algorithm.

a singleton, we find the invariancy interval to the right from it. Now, we have an edge of one of the invariancy regions and we can get an initial inner point of that invariancy region selecting a point on the edge and utilizing Algorithm 2.1. Using that initial inner point, we can identify the first non-trivial invariancy region including all of its edges and vertices as described on pages 40-46, see Figure 2.7(c).

To enumerate all invariancy regions in the bounding box, we use concepts and tools [36, 114] from computational geometry. The algorithm that we are going to present possess some similarities with polygon subdivision of the space and planar graphs. Our algorithm is essentially the subdivision of the bounding box into convex polyhedrons that can be unbounded. First, we introduce the notation and geometric objects used in computational geometry to describe these type of problems. Second, we show how to use those objects to create the complete algorithm for invariancy region enumeration.

The geometric objects involved in the given problem are vertices, edges and cells (faces), see Figure 2.8. Cells correspond to the non-trivial invariancy regions. Edges and vertices are trivial invariancy regions, each edge connects two vertices. It is important to notice that cells can be unbounded if the corresponding invariancy region is unbounded. That is why we need to extend the representation of the vertex to allow incorporating the information that the ver-

tex can represent the virtual endpoint of the unbounded edge if the corresponding cell is unbounded. For instance, edge e_1 on Figure 2.8 is unbounded, so in addition to its first endpoint v_1, we add another virtual endpoint being any point on the edge except v_1. Consequently, each vertex need to be represented not only by its coordinates (x, y), but also by the third coordinate z that indicates if it is a virtual vertex and the corresponding edge is unbounded. Another note to make is that the optimal partition may not be unique for each vertex or edge. First, at every virtual vertex, the optimal partition is the same as on the corresponding edge. Second, we may have situations when the optimal partition is the same on the incident edges and vertices if those are on the same line (edges e_2 and e_7 and vertex v_3 haves the same optimal partition on Figure 2.8).

The data structures that we use for storing the information about vertices, faces and cells are similar to the ones used in many computational geometry algorithms [36] and are described in Chapter 4. Traversing the cell is usually done counter-clockwise and we are going to follow that convention as well. The extension of the standard storage and representation model is that we allow convex polyhedrons (cells) to be unbounded and that we do not require the vertexes to be in general positions (three vertices can be on one line).

To enumerate all invariancy regions we use two queues that store indices of the cells that are already investigated and to be processed. At the start of the algorithm, the first cell enters the to-be-processed queue and the queue of completed cells is empty (c_1 is entering the to-be-processed queue on Figure 2.8). After that, we identify the cell c_1 including all faces and vertices starting from the known edge e_1 and moving counter-clockwise (note that the virtual vertices corresponding to the unbounded edges are not shown on Figure 2.8). Due to the fact that the optimal partition at the edge between the vertices v_1 and v_2 is the same, we are able to identify only the edge e_2 at the moment. Now, when we have identified all the edges incident to the cell c_1 we can add the potential cells corresponding to each of the edges to the to-be-processed queue, so $e_1 \rightarrow \emptyset$ (infeasible), $e_2 \rightarrow c_2$, $e_3 \rightarrow c_3$. So, we add c_2 and c_3 to the to-be-

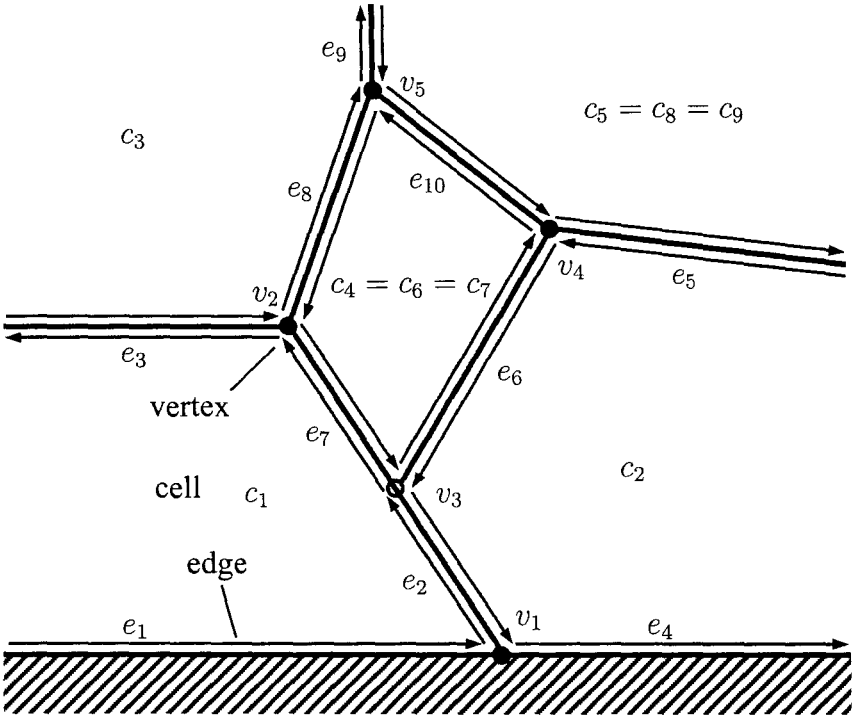


Figure 2.8: Bi-Parametric QO – Computational Geometry Problem Representation.

processed queue and move c_1 to the completed queue. Next, we start processing the cell c_2 as the first element of the to-be-processed queue. At this stage, we identify that the edge e_2 is shorter than the original one and we split it into two edges – e_2 and e_7 (note that the edges e_2, e_7 and vertex v_3 have the same optimal partition). As the result of splitting edge e_2 into two edges, we need to add the cell c_4 that corresponds to the edge e_7 to the to-be-processed queue. We get $e_7 \rightarrow c_4$, $e_2 \rightarrow c_1$ (already processed), $e_4 \rightarrow \emptyset$ (infeasible), $e_5 \rightarrow c_5$, $e_6 \rightarrow c_6$ and add c_4, c_5, c_6 into the to-be-processed queue. Now, c_2 is moved to the completed queue. Next in the to-be processed queue is c_3 that gives us $e_3 \rightarrow c_1$ (already processed), $e_8 \rightarrow c_7$ and $e_9 \rightarrow c_8$. So, c_7 and c_8 are added to

the to-be-processed queue and c_3 is moved to the completed queue. Next, we process c_4 and identify that $c_4 = c_6 = c_7$ based on checking the identified optimal partitions list and identified edges list. Here, $e_{10} \rightarrow c_9$. So, c_4 is moved to the completed queue and c_6, c_7 are removed from the to-be-processed queue. Next in the to-be-processed queue is c_5 and we identify that $c_5 = c_8 = c_9$ and there are no more new edges. As the result, we move c_5 to the completed queue and remove c_8 and c_9 from the to-be-processed queue. The to-be-processed queue is empty now, so we have identified all the invariancy regions.

Algorithm 2.2: Bi-Parametric QO Algorithm for Enumerating All Invariancy Regions.

Data: The QO optimization problem and $\triangle b$, $\triangle c$

Result: Optimal partitions on all invariancy intervals, optimal value function

Initialization: compute bounding box in the "$\epsilon - \lambda$" plane and compute inner point in one of the invariancy regions;

while *not all invariancy regions are enumerated* **do**

  run sub-algorithm to compute all edges and vertices of the current invariancy region;

  add all unexplored regions corresponding to each edge to the to-be-processed queue and move the current region to the queue of completed region indices;

  if *to-be-processed queue of the unexplored regions is not empty* **then**

    pull out the first region from the to-be-processed queue;

    compute an inner point of the new region;

  else

    return the data structure with all the invariancy regions, corresponding optimal partitions and optimal value function;

  end

end

Note that in the description of our algorithm we did not explicitly consider the cases when the optimal partition changes on the open line segments at the boundary of invariancy regions. This situations can potentially happen and our algorithm is able to handle it correctly. We did not encounter these cases on the examples that we have solved. To clarify if the optimal partition is invariant on the open boundary line segments of non-trivial invariancy regions remains for future research.

The proposed Algorithm 2.2 on page 51 runs in linear time in the output size (the constant $C \cdot n$ is 3). But, by the nature of the parametric problem, the number of vertices, edges and faces can be exponential in the input size. In our experiences, the worst case does not happen in practise very often though.

In summary, in this section we have extended the uni-parametric simultaneous perturbation results for QO to the bi-parametric case. Algorithm 2.2 outlined in this section allows identifying all invariancy regions where the optimal partition is invariant by solving a series of uni-parametric QO problems. We can also compute the optimal value function and maximally complementary solutions on each invariancy region.

Even though all presented auxiliary optimization problems can be solved in polynomial time by IPMs and the number of different optimal partitions is finite, enumeration of all invariancy regions may not be achieved in polynomial time due to the fact that the number of different optimal partitions may increase exponentially with the cardinality of the index set. That is why the algorithm presented is linear in the output size, but not in the input size.

2.4 Multi-Parametric Quadratic Optimization

Generalizing the results of bi-parametric QO to the multi-parametric case, when the perturbation parameters appear in the objective function and the constraints are independent, is the natural extension of the previous findings in Section 2.3. Multi-parametric analysis poses many challenges because even extension from

one dimensional parameter space to the two-dimensional parameter space is non-trivial. The complexity of the algorithm grows substantially and the necessity to use a recursive algorithm may appear. The difficulty is that the invariancy regions of the multi-dimensional parameter space are high-dimensional polyhedrons that should be identified and enumerated by an algorithm.

Having efficient techniques and algorithms for bi-parametric QO opens the way of designing algorithms for multi-parametric problems where the parameters λ and/or ϵ are multidimensional vectors. Those technics are mostly application-driven and would probably have a small number of distinct parameters.

The following fundamental properties of the invariancy regions and optimal value function for the bi-parametric QO problem extend to the multi-parametric case (with k parameters):

- the optimal partition $\pi = (\mathcal{B}, \mathcal{N}, \mathcal{T})$ is constant on invariancy regions;
- the invariancy regions $\mathcal{IR}$ are convex sets;
- the invariancy region $\mathcal{IR}$ can be the singleton $\{0\}$, one dimensional set, two dimensional polyhedron, ..., k-dimensional polyhedron that all might be unbounded;
- the optimal value function is a multivariate quadratic function on invariancy region $\mathcal{IR}$;
- the optimal value function is continuous and piecewise multivariate quadratic.

Keeping in mind that fundamental properties of invariancy regions and the optimal value function can be easily extended from the bi-parametric to the multi-parametric case, we can also extend the bi-parametric Algorithm 2.2. The computational challenge of the algorithm would significantly increase, as instead of working in the 2-dimensional space of parameters, we need to explore the k-dimensional space.

To the best of our knowledge, there are no results on optimal partition-based multi-parametric QO in the literature. Properties of the optimal partition-based multi-parametric LO were investigated in [69], but no algorithm was pro-

vided.

An active set-based algorithm for multi-parametric QO was described in [142, 116]. It is designed for solving the following parametric problem with perturbation on the right-hand-side of the constraints only:

$$\begin{array}{ll} \min & \frac{1}{2}x^T Q x \\ \text{s.t.} & Ax \leq b + S\lambda, \end{array} \tag{2.4.1}$$

where $S \in \mathbb{R}^{m \times n}$ and $\lambda \in \mathbb{R}^m$ is the parameter. This kind of problems often appear in model predictive control (see, e.g., [142] for more details and references).

We would like to mention some differences of our algorithmic approach to parametric QO optimization and the algorithm described in [116] which is implemented in [86]. First, in our study we consider simultaneous perturbation in the right-hand-side of the constraints and the linear term of the objective function with different parameters, while in [116] and related publications only perturbation in either the right-hand-side or the linear term of the objective is considered. Second, in [116] the authors define a critical region as the region of parameters where active constraints remain active. As the result, an important precondition for analysis in [116] is the requirement for either making non-degeneracy assumption or exploiting special tools for handling degeneracy, while, our algorithm does not require any non-degeneracy assumptions. Finally, the algorithm for parametric quadratic optimization described in [116] uses a different parameter space exploration strategy than ours. Their recursive algorithm identifies a first critical (invariancy) region, and after that reverses the defining hyperplanes one by one in a systematic process to get a subdivision of the complement set. The regions in the subdivision are explored recursively. As the result, each critical (invariancy) region can be split among many regions and, consequently, all the parts has to be detected. Thus, each of the potentially exponential number of invariancy regions may be split among exponential number of regions, which makes their algorithm computationally expensive.

Chapter 3

Multiobjective and Parametric Optimization

In this chapter we highlight the relationships between multiobjective optimization and parametric optimization that is used to solve such problems. Solution of a multiobjective problem is the set of Pareto efficient points, known in the literature as Pareto efficient frontier or Pareto front. Pareto points can be obtained by using either weighting the objectives or by ε-constrained (hierarchical) method for solving multiobjective optimization models. Using those methods we can formulate them as parametric optimization problems and compute their efficient solution set numerically. We present a methodology that allows tracing the Pareto efficient frontier without discretization of the objective space and without solving the corresponding optimization problem at each discretization point [58].

3.1 Multiobjective Optimization Problems

Let x be an n-dimensional vector of *decision variables*. The multiobjective optimization problem, where the goal is to optimize a number of possibly conflicting objectives simultaneously, is formulated as:

$$\begin{array}{ll} \min & \{f_1(x), f_2(x), \ldots, f_k(x)\} \\ \text{s.t.} & x \in \Omega, \end{array} \tag{3.1.1}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, k$ are (possibly) conflicting objectives and $\Omega \subseteq \mathbb{R}^n$ is a feasible region. Each of the functionsf_i represent an attribute or a *decision criterion* that serves the base for the decision making process.

Multiobjective optimization is a subclass of *vector optimization*, where the vector-valued objective function $f_0 = \{f_1(x), f_2(x), \ldots, f_k(x)\}$ is optimized with respect to a proper convex cone $\mathcal{C}$ which defines preferences. When a vector optimization problem involves the cone $\mathcal{C} = \mathbb{R}_+$, it is known as a *multicriteria* or *multiobjective optimization* problem.

In this chapter we consider *convex multiobjective conic optimization* problems and most of the results hereafter are restricted to that problem class. Moreover, we also mention some of the results available for general multiobjective problems. Problem (3.1.1) is a *convex* multiobjective optimization problem if all the objective functions $f_1, \ldots, f_k$ are convex, and the feasible set Ω is convex as well. For example, it can be defined as $\Omega = \{x : g_j(x) \leq 0,\ h_j(x) = 0\}$, where the inequality constraint functions $g_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, l$ are convex and the equality constraint functions $h_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, m$ are affine. For LO, QO and SOCO problems the set of constraints can be written as $\Omega = \{x : Ax = b,\ x \geq_{\mathcal{K}} 0\}$, where $\mathcal{K}$ is an appropriate convex cone and $Ax = b$ are the equality constraints with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The set Ω, is called the feasible region in the decision space or just the *decision space*.

Definition 3.1.1 *A vector $x^* \in \Omega$ is Pareto optimal (or efficient solution) if there does not exist another $x \in \Omega$ such that $f_i(x) \leq f_i(x^*)$ for all $i = 1, \ldots, k$ and $f_j(x) < f_j(x^*)$ for at least one index j.*

The set of all Pareto optimal (or efficient) solutions $x^* \in \Omega$ is called the Pareto optimal (efficient solution) set Ω_E.

As values of the objective functions are used for making decisions by the decision maker, it is conventional for multiobjective optimization to work in the space of the objective functions, which is called the *objective space*. By mapping

the feasible region into the objective space, we get:

$$Z = \{z \in \mathbb{R}^k \, : \, z = ((f_1(x), f_2(x), \ldots, f_k(x))^T \; \forall \, x \in \Omega)\}.$$

The set Z is the set of objective values of feasible points, it is referred to as the set of *achievable objective values.* Points in the achievable set Z can be ranked into efficient and non-efficient points (see Figure 3.1) that leads to the definition of Pareto optimality.

Analogous definition of Pareto optimality can be stated for an objective vector $z^* \in Z$. Equivalently, z^* is Pareto optimal if the decision vector x^* corresponding to it is Pareto optimal [102].

Definition 3.1.2 *For a given multiobjective problem (3.1.1) and Pareto optimal set Ω_E, the Pareto front is defined as:*

$$Z_N = \{z^* = (f_1(x^*), \ldots, f_k(x^*))^T \mid x^* \in \Omega_E\}.$$

A set Z_N of Pareto optimal (also called nondominated or efficient) solutions z^* forms the *Pareto efficient frontier* or *Pareto front.* The Pareto front, if $k = 2$, is also known as the *optimal trade-off curve* and for $k > 2$ it is called the *optimal trade-off surface* or the *Pareto efficient surface.*

Solution methods are designed to help the decision maker to identify and choose a point on the Pareto front. Identifying the whole frontier is computationally challenging, and often it cannot be performed in reasonable time. Solution methods for multiobjective optimization are divided into the following categories [102]:

- *a priori methods* are applied when the decision maker's preferences are known a priory; those include the value function method, lexicographic ordering and goal programming.
- *iterative methods* guide the decision maker to identify a new Pareto point from an existing one (or existing multiple points), the process is stopped when the decision maker is satisfied with the actual efficient point.

- *a posteriori methods* are used to compute the Pareto front or some of its parts; those methods are based on the idea of scalarization, namely transforming the multiobjective optimization problem into a series of single-objective problems; a posteriori methods include *weighting methods*, the *ε-constrained method* and related scalarization techniques.

Computing the Pareto front can be challenging as it does not posses known structure in most of the cases, and, consequently, discretization in the objective space is frequently used to compute it. The problem is that discretization is computationally costly in higher dimensions, and discretization is not guaranteed to produce all the (or desired) points on the Pareto front.

It turns out that for some classes of multiobjective optimization problems the structure of the efficient frontier can be identified. Those include multiobjective LO, QO and SOCO optimization problems. For those classes of problems, the Pareto efficient frontier can be sub-divided into pieces (subsets) that have specific properties. These properties allow the identification of each subsets of the frontier. The piece-wise structure of the Pareto front also provides additional information for the decision maker.

Before looking at the scalarization solution techniques for multiobjective optimization, that allow us to identify all nondominated (Pareto efficient) solutions, we need to introduce a number of concepts and some theoretical results.

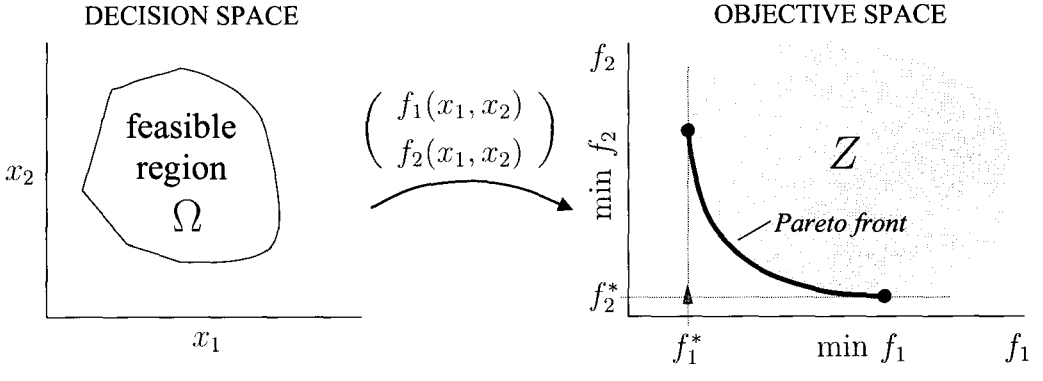


Figure 3.1: Mapping the Decision Space into the Objective Space.

Definition 3.1.3 *An objective vector $z^* \in Z$ is weakly Pareto optimal if there does not exist another decision vector $z \in Z$ such that $z_i < z_i^*$ for all $i = 1, \ldots, k$.*

The set of weakly Pareto efficient (nondominated) vectors is denoted by Z_{wN}. It follows that $Z_N \subseteq Z_{wN}$. When unbounded trade-offs between objectives are not allowed, Pareto optimal solutions are called *proper* [102]. The set of properly efficient vectors is denoted as Z_{pN}.

Both sets Z_{wN} (weak Pareto front) and Z_N (Pareto front) are connected if the functions f_i are convex and the set Ω satisfies one of the following properties [44]:

- Ω is a compact, convex set;
- Ω is a closed, convex set and $\forall z \in Z$, $\Omega(z) = \{x \in \Omega : f(x) \leq z\}$ is compact.

Let us denote by $\mathbb{R}^k_+ = \{z \in \mathbb{R}^k : z \geq 0\}$ the nonnegative orthant of $\mathbb{R}^k$. Consider the set:

$$\mathcal{A} = Z + \mathbb{R}^k_+ = \{z \in \mathbb{R}^k : f_i(x) \leq z_i,\ i = 1, \ldots, k,\ x \in \Omega\},$$

that consists of all values that are worse than or equal to some achievable objective value. While the set Z of achievable objective values need not be convex, the set $\mathcal{A}$ is convex, when the multiobjective problem is convex [15].

Definition 3.1.4 *A set $Z \in \mathbb{R}^k$ is called $\mathbb{R}^k_+$-convex if $Z + \mathbb{R}^k_+$ is convex.*

A point $x \in \mathcal{C}$ is a minimal element with respect to componentwise inequality induced by $\mathbb{R}^k_+$ if and only if $(x - \mathbb{R}^k_+) \cap \mathcal{C} = x$. The minimal elements of $\mathcal{A}$ are exactly the same as the minimal elements of the set Z. This also means that any hyperplane tangent to the Pareto efficient surface is a supporting hyperplane – the Pareto front is on one side of the hyperplane [41]. It follows that the Pareto front must belong to the boundary of Z [41].

Proposition 3.1.1 $Z_N = \left(Z + \mathbb{R}^k_+\right)_N \subset \mathrm{bd}(Z)$.

When talking about convex multiobjective optimization problems, it is useful to think of the Pareto front as a function, and not as a set. Under assumptions about convexity of the functions f_i and the set Ω for bi-objective optimization problems ($k = 2$), the (weakly) Pareto front is a convex function [134]. Unfortunately, when $k > 2$ it is not the case even for linear multiobjective optimization problems.

Most a posteriori methods for solving multiobjective optimization problems are based on scalarization techniques. Let us consider the two most popular scalarization methods:

- weighting method;
- ε-constraint method.

3.1.1 Weighting Method

The idea of the weighting method is to assign weights to each objective function and optimize the weighted sum of the objectives. A multiobjective optimization problem can be solved with the use of the *weighting method* by optimizing single-objective problems of the type

$$\begin{aligned} \min \quad & \sum_{i=1}^{k} w_i f_i(x) \\ \text{s.t} \quad & x \in \Omega, \end{aligned} \tag{3.1.2}$$

where f_i is linear, convex quadratic or second order conic function in our case, $\Omega \subseteq \mathbb{R}^n$ (convex), $w_i \in \mathbb{R}$ is the weight of the i-th objective, $w_i \geq 0$, $\forall i = 1, \ldots, k$ and $\sum_{i=1}^{k} w_i = 1$. Weights w_i define the importance of each objectives. Due to the fact that each objectives can be measured in different units, the objectives may have different magnitudes. Consequently, for the weight to define the relative importance of objectives, all objectives should be normalized first. Some of the normalization methods are discussed in [63]. As we intend to compute the whole Pareto front, normalization is not required.

It is known that the weighting method produces weakly efficient solutions when $w_i \geq 0$ and efficient solutions if $w_i > 0$ for all $i = 1, \ldots, k$ [102]. For

convex multiobjective optimization problems any Pareto optimal solution x^* can be found by the weighting method.

Let us denote by $\mathcal{S}(w, Z) = \{\hat{z} \in Z : \hat{z} = \text{argmin}_{z \in Z} w^T z\}$ the set of optimal points of Z with respect to w. In addition, we define

$$\mathcal{S}(Z) = \bigcup_{w>0,\ \sum_{i=1}^{k} w_i=1} \mathcal{S}(w, Z), \qquad \mathcal{S}_0(Z) = \bigcup_{w\geq 0,\ \sum_{i=1}^{k} w_i=1} \mathcal{S}(w, Z).$$

As Z is $\mathbb{R}^k_+$-convex set in our case, we get [41]:

$$\mathcal{S}(Z) = Z_{pN} \subset Z_N \subset \mathcal{S}_0(Z) = Z_{wN}. \tag{3.1.3}$$

In addition, if $\hat{z}$ is the unique element of $\mathcal{S}(w, Z)$ for some $w \geq 0$, then $\hat{z} \in Z_N$ [41]. The last observation combined with (3.1.3), allows us identifying the whole (weak) Pareto front with the use of the weighting method.

3.1.2 ε-Constrained Method

For illustration purposes, we first consider a problem with two objective functions. Multiobjective optimization can be based on ranking the objective functions in descending order of importance. Each objective function is then minimized individually subject to a set of additional constraints that do not allow the values of each of the higher ranked functions to exceed a prescribed fraction of their optimal values obtained in the previous step. Suppose that f_2 has higher rank than f_1. We then solve

$$\min \{f_2(x) : x \in \Omega\},$$

to find the optimal objective value f_2^*. Next, we solve the problem

$$\begin{array}{ll} \min & f_1(x) \\ \text{s.t.} & f_2(x) \leq (1+\epsilon) f_2^*, \\ & x \in \Omega. \end{array}$$

Intuitively, the hierarchical ranking method can be thought as saying "f_2 is more important than f_1 and we do not want to sacrifice more than ϵ percentage of the optimal value of f_2 to improve f_1."

Considering the general case of k objective functions and denoting the right-hand-side term of the constraints on the objective functions' values by $\varepsilon_j = (1+\epsilon_j)f_j^*$, we get the following single-objective optimization problem, which is known as the *ε-constrained method*:

$$(MOC_\varepsilon) \qquad \begin{array}{ll} \min & f_\ell(x) \\ \text{s.t} & f_j(x) \leq \varepsilon_j,\ j = 1,\ldots,k,\ j \neq \ell \\ & x \in \Omega. \end{array} \tag{3.1.4}$$

Every solution x^* of the ε-constrained problem (3.1.4) is weakly Pareto optimal [102], so formulation (3.1.4) can be used to compute weak Pareto front Z_{wN}.

Let x^* solve (3.1.4) with $\varepsilon_j^* = f_j(x^*)$, $j \neq \ell$. Then x^* is *Pareto optimal* [23, 42] if:

1) x^* solves (3.1.4) for every $\ell = 1,\ldots,k$;
2) x^* is the unique solution of (3.1.4);
3) Lin's conditions [89, 90].

The third set of necessary and sufficient conditions for (strong) Pareto optimality of optimal solutions is described in [23] based on the results of Lin [89, 90]. Let us define

$$\phi_\ell(\varepsilon) = \min\{f_\ell(x) : x \in \Omega,\ f_j(x) \leq \varepsilon_j \text{ for each } j \neq \ell\}.$$

The following theorem [23] establishes that x^* is Pareto optimal if the optimal value of (MOC_{ε^0}) is strictly greater than $f_\ell(x^*)$ for any $\varepsilon^0 \leq \varepsilon^*$.

Theorem 3.1.1 *Let x^* solve (3.1.4) with $\varepsilon_j^* = f_j(x^*)$, $j \neq \ell$. Then x^* is Pareto optimal solution if and only if $\phi_\ell(\varepsilon) > \phi_\ell(\varepsilon^*)$ for all ε such that $\varepsilon \leq \varepsilon^*$ and for each ε (3.1.4) has an optimal solution with finite optimal value.*

In many cases, conditions 2) and 3) can be verified to identify the Pareto front Z_N. For instance, the second condition holds when all the objective functions $f_j(x)$ are strictly convex. Condition 3) can be verified if function $\phi_\ell(\varepsilon)$ is computed by parametric optimization techniques, see Section 3.2.

3.2 Multiobjective Optimization via Parametric Optimization

By now, the reader may have understood that multiobjective optimization problems are closely related to, and can be represented as parametric optimization problems. Consequently, we may use algorithms of parametric optimization to solve multiobjective optimization problems and to compute the Pareto fronts. Before defining the relations between multiobjective optimization and parametric optimization more formally, we mention that multiobjective LO, QO and, to some extent, SOCO problems can be efficiently solved by parametric optimization algorithms. Parametric optimization techniques exist for wider classes of problems, but computational complexity may prevent using those directly to identify efficient frontiers.

The main idea of this chapter is that we can solve multiobjective optimization problems using parametric optimization techniques. *A posteriori multiobjective optimization techniques* are based on parameterizing (scalarizing) the objective space and solving the resulting parametric problem. Consequently, parametric optimization algorithms can be utilized to solve multiobjective optimization problems.

Based on the weighting method (3.1.2) and choosing the vector of weights as $w = (\lambda_1, \ldots, \lambda_{k-1}, 1)^T \geq 0$, as w can be scaled by a positive constant, for the weighted objective function $\sum_i w_i f_i(x)$, we can formulate the parametric optimization problem with the λ_i parameters in the objective function as

$$\begin{array}{rcll} \phi(\lambda_1, \ldots, \lambda_{k-1}) & = & \min & \lambda_1 f_1(x) + \ldots + \lambda_{k-1} f_{k-1}(x) + f_k(x) \\ & & \text{s.t.} & x \in \Omega, \end{array} \tag{3.2.1}$$

for computing weakly Pareto optimal solutions, or $(\lambda_1, \ldots, \lambda_{k-1})^T > 0$ for computing Pareto optimal solutions. Formulation (3.2.1) is known as the *Lagrangian problem* [23] and possesses almost identical properties as the weighting problem (3.1.2).

Based on the ε-constrained method (3.1.4) we can present the following

parametric problem:

$$\begin{array}{rcll} \phi(\varepsilon_1, \ldots, \varepsilon_{k-1}) & = & \min & f_k(x) \\ & & \text{s.t} & f_i(x) \leq \varepsilon_i, \ i = 1, \ldots, k-1 \\ & & & x \in \Omega, \end{array} \tag{3.2.2}$$

where $\varepsilon_1, \ldots, \varepsilon_{k-1}$ are parameters in the right-hand-side of the constraints. In this case, the optimal value function $\phi(\varepsilon_1, ..., \varepsilon_{k-1})$ includes the Pareto front as a subset.

It is not hard to observe that the parametric problems (3.2.1) and (3.2.2) are equivalent to (3.1.2) and (3.1.4), respectively, but they are just written in the forms used in the parametric optimization literature. The relationships between those formulations and their properties are extensively studied in [23].

Algorithms and techniques developed for solving parametric optimization problems are described in Chapters 2 and 8. Note that the optimal value function $\phi(\varepsilon)$ of problem (3.2.2) is the boundary of the set $\mathcal{A}$ and the Pareto front is a subset of that boundary. These results are illustrated by examples in Chapter 7.

As we learned in this section, multiobjective optimization problems can be formulated as parametric optimization problems. Some classes of multiobjective optimization problems that include linear and convex quadratic optimization problems can be efficiently solved using parametric optimization algorithms. Parametric optimization allows not only computing Pareto efficient frontiers (surfaces), but also identifying piece-wise structures of those frontiers. Structural description of Pareto fronts gives functional form of each of its pieces and thus helps decision makers to make better decisions.

3.3 Multiobjective and Parametric Quadratic Optimization

Results described in Sections 3.1 and 3.2 apply to general convex multiobjective optimization problems. In contrast, parametric optimization techniques discussed in this thesis apply to LO, QO and SOCO problems only. In this

section we specialize the formulations presented in Section 3.2 to the parametric optimization problem classes described in Chapter 2 and 8.

We define the *multiobjective quadratic optimization* problem as a convex multiobjective problem with one convex quadratic objective function f_k and $k-1$ linear objectives $f_1, \ldots, f_{k-1}$ subject to linear constraints. For the multiobjective QO problem the weighted sum formulation (3.2.1) specializes to

$$\begin{array}{rcll} \phi(\lambda_1, \ldots, \lambda_{k-1}) & = & \min & \lambda_1 c_1^T x + \ldots + \lambda_{k-1} c_{k-1}^T x + \frac{1}{2} x^T Q x \\ & & \text{s.t.} & Ax = b \\ & & & x \geq 0, \end{array} \tag{3.3.1}$$

and the ε-constrained formulation (3.2.2) becomes

$$\begin{array}{rcll} \phi(\varepsilon_1, \ldots, \varepsilon_{k-1}) & = & \min & \frac{1}{2} x^T Q x \\ & & \text{s.t.} & c_i^T x \leq \varepsilon_i, \ i = 1, \ldots, k-1 \\ & & & Ax = b \\ & & & x \geq 0. \end{array} \tag{3.3.2}$$

Parametric QO formulations (3.3.1) and (3.3.2) can be solved with algorithms developed in Chapter 2. The uni-parametric case corresponds to an optimization problem with two objectives. A bi-parametric QO algorithm allows solving multiobjective QO problems with three objectives. Multiobjective problems with more than three objectives require multi-parametric optimization techniques. Note that in formulations (3.3.1) and (3.3.2), parameters appear in the objective function and in the right-hand side of the constraints, respectively.

Multiobjective QO problems are historically solved by techniques that approximate the Pareto front [52, 51]. An alternative approach is the parametric optimization discussed in this thesis. Examples of multiobjective QO problems appearing in finance are solved with parametric QO techniques in Chapter 7.

If we allow for more than one convex quadratic objective in the multiobjective optimization problem, formulations (3.3.1) and (3.3.2) become parametric QOCO. It happens due to the fact that now quadratic functions appear in the constraints as well. Parametric SOCO, that includes parametric QCQO

problems, is a more general class of problems. Preliminary results for solving parametric SOCO problems are described in Chapter 8. Properties of multiobjective optimization problems with more than one convex quadratic objectives and linear constraints are discussed in [59].

Chapter 4

Implementation of Parametric Optimization

Implementation details of the parametric LO and QO package are described in this chapter. In Section 4.1 we provide a number of simple numerical examples to illustrate parametric optimization concepts and show the desired output of the parametric solver. Implementation ideas and results discussed in Section 4.2 are extended from [123].

4.1 Illustrative Examples

Here we present a number of illustrative numerical examples that show the output of the parametric solver that is based on the results described in Chapters 2. Computations related to finding optimal solutions of auxiliary subproblems can be performed by using any IPM solver for LO and QO problems due to the fact that IPMs find a maximally complementary solution in the limit. We have used the MOSEK [105] solver for our computations when solving the sub-problems. More complex examples appearing in practical problems are solved in Chapters 5 and 7.

4.1.1 Uni-Parametric QO

Let us consider the following convex QO problem with $x, c \in \mathbb{R}^5$, $b \in \mathbb{R}^3$, $Q \in \mathbb{R}^{5\times 5}$ being a positive semidefinite symmetric matrix, $A \in \mathbb{R}^{3\times 5}$ with $\text{rank}(A) = 3$. The problem data is

$$Q = \begin{bmatrix} 4 & 2 & 0 & 0 & 0 \\ 2 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad c = \begin{bmatrix} -16 \\ -20 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \triangle c = \begin{bmatrix} 7 \\ 6 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$A = \begin{bmatrix} 2 & 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 2 & 5 & 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 11 \\ 8 \\ 20 \end{bmatrix}, \quad \triangle b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

With this data the perturbed convex QO problem (QP_λ) is

$$\begin{array}{lll} \min & (-16+7\lambda)x_1 + (-20+6\lambda)x_2 + 2x_1^2 + 2x_1x_2 + \frac{5}{2}x_2^2 & \\ \text{s.t.} & 2x_1 + 2x_2 + x_3 = 11+\lambda & \\ & 2x_1 + x_2 + x_4 = 8+\lambda & (4.1.1) \\ & 2x_1 + 5x_2 + x_5 = 20+\lambda & \\ & x_1,\ x_2,\ x_3,\ x_4,\ x_5 \geq 0. & \end{array}$$

Table 4.1: Transition Points, Invariancy Intervals, and Optimal Partitions for the Illustrative Uni-Parametric QO Problem.

Inv. Intervals and Tr. Points	$\mathcal{B}$	$\mathcal{N}$	$\mathcal{T}$	$\phi(\lambda)$
$\lambda = -8.0$	{3,5}	{1,4}	{2}	
$-8.0 < \lambda < -5.0$	{2,3,5}	{1,4}	$\emptyset$	$68.0\lambda + 8.5\lambda^2$
$\lambda = -5.0$	{2}	{1,3,4,5}	$\emptyset$	
$-5.0 < \lambda < 0.0$	{1,2}	{3,4,5}	$\emptyset$	$-50.0 + 35.5\lambda + 4\lambda^2$
$\lambda = 0.0$	{1,2}	$\emptyset$	{3,4,5}	
$0.0 < \lambda < 1.739$	{1,2,3,4,5}	$\emptyset$	$\emptyset$	$-50.0 + 35.5\lambda - 6.9\lambda^2$
$\lambda = 1.739$	{2,3,4,5}	$\emptyset$	{1}	
$1.739 < \lambda < 3.333$	{2,3,4,5}	{1}	$\emptyset$	$-40.0 + 24.0\lambda - 3.6\lambda^2$
$\lambda = 3.333$	{3,4,5}	{1}	{2}	
$3.333 < \lambda < +\infty$	{3,4,5}	{1,2}	$\emptyset$	0

The computational results we are interested in are presented in Table 4.1. The set Λ for the optimal value function $\phi(\lambda)$ is $[-8, +\infty)$. Figure 4.1 depicts the

graph of $\phi(\lambda)$. Transition points and the optimal partitions at each transition point and on the invariancy intervals are identified by applying Algorithm 2.1. The optimal value function on the invariancy intervals is computed by using Theorem 2.2.4. Convexity, concavity or linearity of the optimal value function can be determined by the sign of the quadratic term of the optimal value function (see Table 4.1). As shown in Figure 4.1, the optimal value function is convex on the first two invariancy intervals, concave on the third and fourth and linear on the last one. The first order derivative does not exists at the transition point $\lambda = -5$.

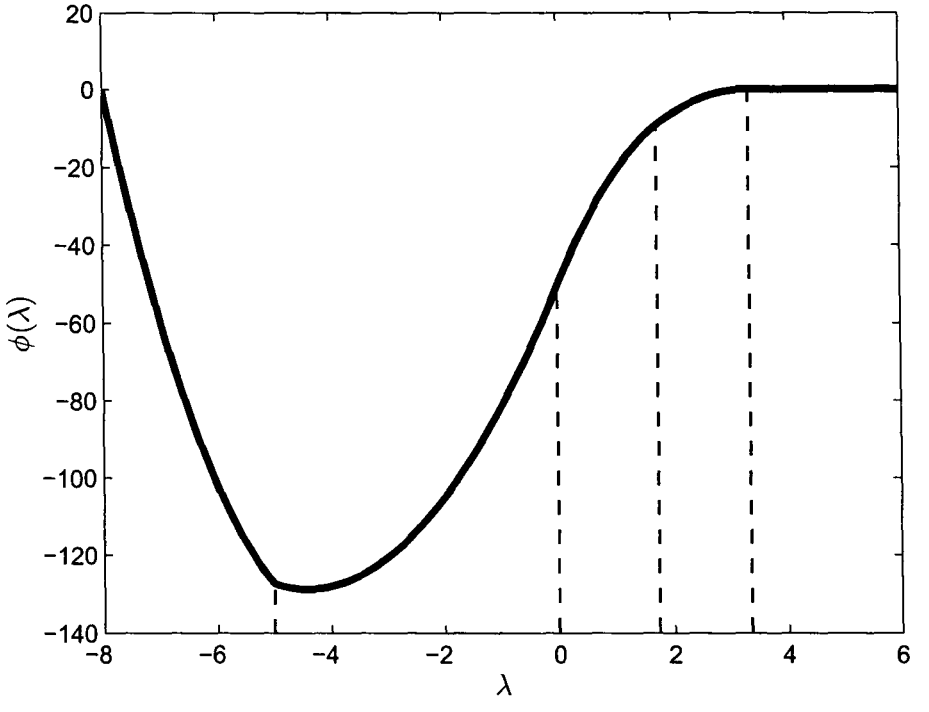


Figure 4.1: The Optimal Value Function for the Illustrative Uni-Parametric QO Problem.

4.1.2 Bi-Parametric QO

Here we present illustrative numerical results for a simple bi-parametric QO example by using Algorithm 2.2 as outlined in Section 2.3. Let us consider the following QO problem with $x, c \in \mathbb{R}^5$, $b \in \mathbb{R}^3$, $Q \in \mathbb{R}^{5\times 5}$ being a positive semidefinite symmetric matrix, $A \in \mathbb{R}^{3\times 5}$ with $\text{rank}(A) = 3$. The problem data is as follows

$$Q = \begin{bmatrix} 4 & 2 & 0 & 0 & 0 \\ 2 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \; c = \begin{bmatrix} -16 \\ -20 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \; \triangle c = \begin{bmatrix} 7 \\ 6 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$A = \begin{bmatrix} 2 & 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 2 & 5 & 0 & 0 & 1 \end{bmatrix}, \; b = \begin{bmatrix} 11 \\ 8 \\ 20 \end{bmatrix}, \; \triangle b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

With this data the perturbed QO instance is

$$\begin{array}{lll} \min & (-16+7\lambda)x_1 + (-20+6\lambda)x_2 + 2x_1^2 + 2x_1x_2 + \frac{5}{2}x_2^2 & \\ \text{s.t.} & 2x_1 + 2x_2 + x_3 = 11+\epsilon & \\ & 2x_1 + x_2 + x_4 = 8+\epsilon & (4.1.2) \\ & 2x_1 + 5x_2 + x_5 = 20+\epsilon & \\ & x_1,\, x_2,\, x_3,\, x_4,\, x_5 \geq 0. & \end{array}$$

The result of our computations is presented in Figure 4.2 and Figure 4.3. Figure 4.2 shows the invariancy regions, the corresponding optimal partitions and the equations for the optimal value function. The optimal partitions for the invariancy intervals are shown in ovals, where each letter corresponds to the corresponding index being in one of the tri-partition sets $\mathcal{B}$, $\mathcal{N}$ or $\mathcal{T}$. The partitions for the transition points are shown next to them. The solid dots correspond to the cases where the optimal partition in those transition points are different from the partitions on the neighboring invariancy intervals and invariancy regions. The circle at the point $\lambda = 10/3$, $\epsilon = -8$ corresponds to the case when the optimal partition for the whole line $\epsilon = -8$ is the same, but it differs for the other segments ending at the point. The graph of $\phi(\epsilon, \lambda)$ and the corresponding invariancy regions are presented in Figure 4.3. The piecewise

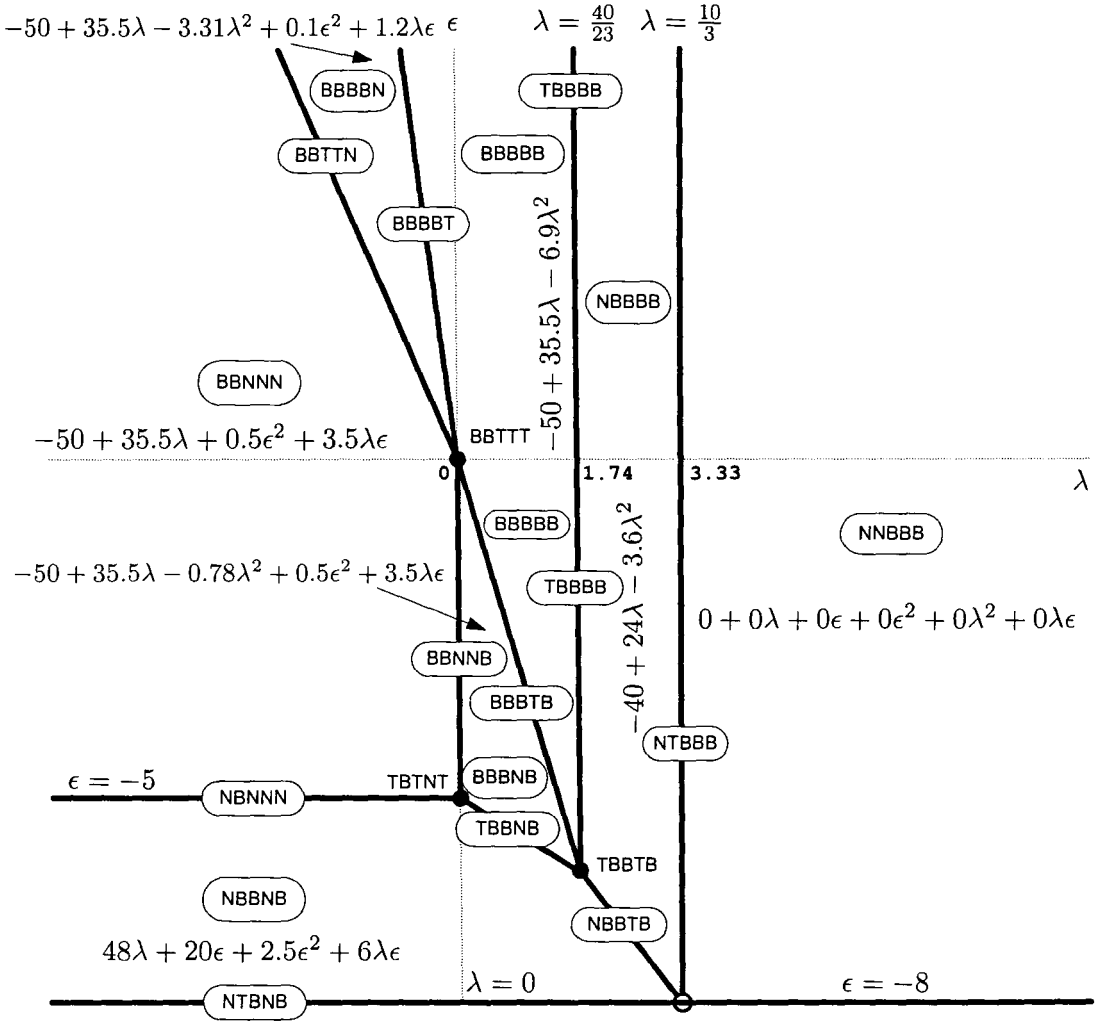


Figure 4.2: The Optimal Partitions and the Invariancy Regions for Bi-Parametric QO Illustrative Example.

quadratic optimal value function is drawn in different colors that correspond to the invariancy regions.

4.2 Implementing the Parametric Algorithm for LO and QO

Implementing parametric optimization into optimization software packages remains one of the challenges. Unfortunately, available software for parametric optimization is very limited. Commercial optimization packages such as CPLEX [74] and MOSEK [105] include basic sensitivity analysis for LO that is based on an optimal basis. In addition, MOSEK provides optimal partition based sensitivity analysis for LO. As parametric optimization is the generalization of sensitivity analysis, techniques for identifying invariancy and stability regions have to be implemented on the top of sensitivity analysis available in those packages. Experimentation with active-set based multi-parametric optimization for LO and QO can be performed with MPT (Multi-Parametric Toolbox for MATLAB) [85]. The MPT toolbox can be called from the YALMIP modeling environment [92].

The goal of the parametric optimization from the algorithmic point of view is to calculate the optimal solution for all relevant values of the parameters without using discretization approaches that are non-rigorous (since there is no guarantee for optimality between the mesh points). Therefore, algorithms for parametric optimization divide the parameter space into regions of optimality; for each region infeasibility is established or an optimal solution is given as a smooth function of the parameters for that region.

Theoretical developments presented in Chapter 2 allow us to develop a computational algorithm for solving bi-parametric QO problems. All the subproblems can be efficiently solved with an IPM. In this section we discuss data structures, implementation details and numerical difficulties that arise when implementing the bi-parametric QO algorithm. The prototype for our implementation is Algorithm 2.2 and we used MATLAB as our software environment.

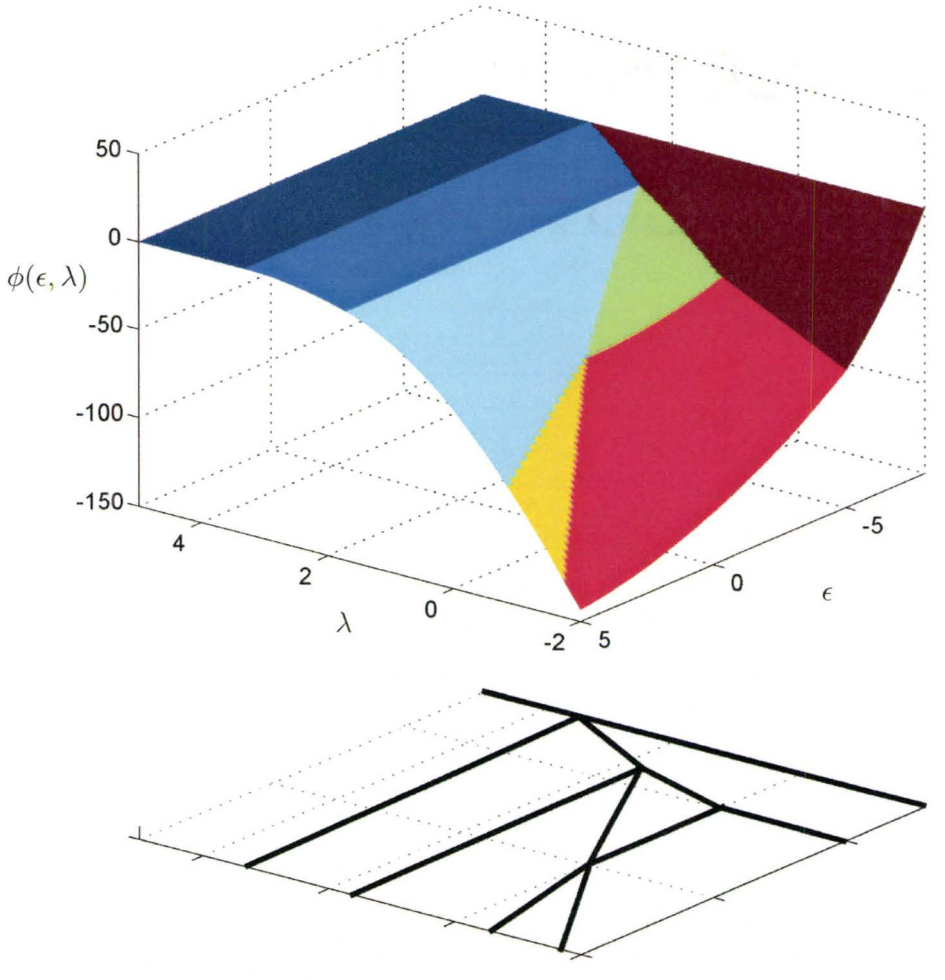


Figure 4.3: The Optimal Value Function for Bi-Parametric QO Illustrative Example. Note that the image is rotated for better visibility.

Data Structures

The data structures that we use for storing the information about vertices, faces and cells are similar to the ones used in many computational geometry algorithms [36]. Traversing the cell is usually done counter-clockwise and we are going to follow that convention as well. The extension of the standard storage and representation model is that we allow convex polyhedrons (cells) to be unbounded and that we do not require the vertexes to be in general positions (three vertexes can be on one line). Moreover, we add some extra fields to the records representing each geometric object. The structures of the records corresponding to vertices, edges, cells and optimal partitions are:

```
vertex {
    vertex id
    coordinates (x,y,z)
    optimal partition code
    }
edge {
    edge id
    vertex_1 id
    vertex_2 id
    incident cell_1
    incident cell_2
    optimal partition code
}
cell {
    cell id
    list of edges
    optimal partition code
}
optimal partition {
    optimal partition code
    list of objects with this optimal partition
}
```

Optimal partitions are numerically encoded as an integer number for minimizing the storage as follows: partition $\mathcal{BB}\ldots\mathcal{BB} = 0$, $\mathcal{BB}\ldots\mathcal{BN} = 1$, $\mathcal{BB}\ldots$

$\mathcal{BT} = 2$, $\mathcal{BB}\ldots\mathcal{NB} = 3$, etc. To do the conversion we apply the summation formula $\sum_{i=0}^{n-1} d_i 3^i$, where $d_i = 0$ if $i \in \mathcal{B}$, $d_i = 1$ if $i \in \mathcal{N}$ and $d_i = 2$ if $i \in \mathcal{T}$. This encoding (optimal partition code) allows not only saving storage, but also to establish the lexicographic order of the identified optimal partitions and to use binary search tree for verifying if the identified optimal partition was already encountered. We are using binary search trees for searching quickly among identified vertices, edges, cells and optimal partitions.

Determining Optimal Partitions and Support Sets

Determination of the optimal partition for a given maximally complementary optimal solution or determination of the support set for a given optimal solution is a challenging task primarily due to numerical reasons. Basic ideas about numerical determination of the optimal partition are discussed in [123] and we recall those here. From Chapter 2 we know that for a given maximally complementary solution (x^*, y^*, s^*):

$$\begin{aligned} i \in \mathcal{B} \quad &\text{if} \quad s_i^* = 0 \text{ and } x_i^* > 0, \\ i \in \mathcal{N} \quad &\text{if} \quad x_i^* = 0 \text{ and } s_i^* > 0, \\ i \in \mathcal{T} \quad &\text{if} \quad x_i^* = 0 \text{ and } s_i^* = 0. \end{aligned}$$

Unfortunately, numerical solution produced by a LO/QO solver may not allow to determine the optimal partition or support set in 100% of the cases. So, we introduce a zero tolerance parameter `tol_zero` (the default value is 10^{-5}, which performs quite well in practice), and compare the entries of the vectors x^* and s^* to it. As a result, we adopt the following strategy for determining the optimal partition (support set):

$$\begin{aligned} \text{if} \quad & x_i \leq \texttt{tol\_zero} \text{ and } s_i \leq \texttt{tol\_zero} && \text{then} \quad i \in \mathcal{T} \\ \text{elseif} \quad & x_i > \texttt{tol\_zero} \text{ and } s_i < \texttt{tol\_zero} && \text{then} \quad i \in \mathcal{B} \\ \text{elseif} \quad & x_i < \texttt{tol\_zero} \text{ and } s_i > \texttt{tol\_zero} && \text{then} \quad i \in \mathcal{N} \\ \text{elseif} \quad & x_i \geq s_i && \text{then} \quad i \in \mathcal{B} \\ \text{else} \quad & && i \in \mathcal{N} \end{aligned}$$

The methodology described above does not give the desired results in all cases. Even in the linear case when the partition consists of two sets $\mathcal{B}$ and $\mathcal{N}$ only, the task is not easy. As El-Bakry et al. [45] pointed out, reliable identification of the optimal partition may require computation with 16 or more digits of precision. In the quadratic case the tri-partition introduces even more complications as there are 3 sets and the differences between the entries may be even smaller.

Theoretically, optimal partition identification can be performed by solving a number of LO/QO sub-problems as described in [150]. Additional ideas about identifying the optimal partition can be also found in [11, 73]. We left numerical testing of those techniques as future work.

Implementation Details

The implementation of the computational algorithm contains some complications that are worth to mention. The interested reader can find more details about it in Romanko [123]. First, due to numerical errors the determination of the optimal partition and a maximally complementary optimal solution, or the determination of the support set for a given optimal solution is a troublesome task. In contrast with the theoretical results, the numerical solution produced by a QO solver may not allow to determine the optimal partition or support set with 100% reliability. Introducing a zero tolerance parameter and using heuristics described on page 75 may improve the situation. For problems with hundreds or thousands of variables, the probability of getting one or more "problematic" coordinates is very high. Wrongly determined tri-partition may lead to an incorrect invariancy interval, if any. The situation can be improved by resolving the problem for another parameter value close to the current one. Another possibility to overcome this difficulty in implementation is to resolve the problem with fixed "non-problematic" coordinates in order to obtain a more precise solution for the problematic ones.

Second, incorrectly determined optimal partition or support sets, as well as

numerical difficulties, may prevent one of the auxiliary subproblems to be solved, that can be used as an indicator of incorrectly identified partitions. In this case, we restart the algorithm from a parameter value sufficiently close to the current one in order to get the solutions for the whole interval or region.

Third, when identifying boundaries of invariancy regions, numerical sensitivities may prevent getting those correctly. Consider the case when the problem (2.3.22) is solved for two distinct nonzero values t_1 and t_2 and two points $(\epsilon(t_1), \lambda(t_1))$ and $(\epsilon(t_2), \lambda(t_1))$ of the invariancy region boundary having the same computed optimal partition. In that case, Theorem 2.3.7 allows identifying the two vertices of the invariancy region. If the two points $(\epsilon(t_1), \lambda(t_1))$ and $(\epsilon(t_2), \lambda(t_1))$ are computed with (small) numerical errors, identifying the boundary becomes challenging as the optimal partition stays constant between these two points only theoretically.

The prototype parametric package for solving bi-parametric QO problems is implemented in MATLAB. Its structure is similar to the uni-parametric QO package developed in [123], and thus we refer the interested reader to consult flowcharts and interface details there. The output produced by the parametric package is summarized by the data structures described on page 74.

Computational Tests and Analysis of Results

In this section we discuss the performance of our parametric optimization package on a set of parametric QO problems solved in Chapters 5 and 7. As there are no other solvers that can perform completely analogous analysis, and the parametric problem solution time is predetermined by the solution time of the auxiliary problems and the number of invariancy regions, we do not aim to showcase and compare the performance of our software with respect to solution time. That is why we only present the output of our computational results on selected QO/LO problems for illustration purposes. All computations are performed on a Linux server with 8 x Opteron 885 CPUs (16 cores, but jobs run on 1 core) and 64 GB RAM.

Small and medium-size problems are handled pretty well by our prototype implementation. For the large-size parametric problems (number of variables $n \geq 1000$), numerical difficulties occur often, especially due to problems with determination of optimal partitions and support sets.

Summarizing shortly the computational experience of the parametric package the following conclusions can be drawn.

- The performance of the parametric quadratic solver depends on the number of variables. For small- and medium-size problems it performs well, but large-size problems represent a significant challenge. It mostly happens because of numerical troubles occurring when solving the auxiliary subproblems, or due to difficulties with determination of the optimal partition or support sets.
- Robustness of the parametric package is an important issue. It was designed to handle unexpected situations. The package tries to find the invariancy regions, optimal value function, etc. on the whole parameter space. The software makes attempts to recover when some of the auxiliary subproblems are not solved by the QO/LO solver.
- Difficulties in determining the optimal partition can be mitigated by the strategy described on page 75. In addition, we can use other strategies for determining the optimal partition, such as doing heuristic analysis of the problematic indices, or solving a reduced parametric problem that is obtained by eliminating those primal and dual variables that are determined to be in either $\mathcal{B}$ or $\mathcal{N}$.
- The QO/LO solver MOSEK which is currently used for solving the auxiliary subproblems performs quite well.
- Warm-start strategies can be potentially used to reduce the solution time. As for now, IPMs warm-start strategies for LO and QO provide only small reduction in computation time [81, 61, 115]. Due to that reason, warm-start strategies for parametric optimization were not investigated during this research.

Chapter 5

Selected Applications

In this chapter we present non-financial examples of multiobjective optimization problems that can be formulated and solved via parametric optimization. Those examples and their corresponding parametric formulations are described in the following sections.

5.1 Intensity Modulated Radiation Therapy Treatment Planning

Cancer is one of the leading causes of death and is claiming about 75,000 lives in Canada every year. More than 50% of cancer patients in Canada receive radiation therapy treatment. Radiation therapy is used for cancer treatment due to the fact that tumor cells are more susceptible to ionizing radiation than healthy cells. As a result, cancerous cells are more likely to die during the treatment, while healthy tissues have higher chances of survival. To spare healthy tissues surrounding a tumor, that include skin and other organs at risk, shaped radiation beams are administered from several angles and intersecting at the tumor. Consequently, the tumor gets significantly larger absorbed dose than the surrounding healthy tissues.

Intensity modulated radiation therapy (IMRT) allows modulating radiation intensity across the beam and conforming the treatment volume to a tumor

shape. Linear X-ray accelerators with multileaf collimators are used in IMRT to deliver precise radiation doses to tumors. Treatment planning and optimization allow delivering the radiation dose which is consistent with the 3D shape of the tumor by controlling (modulating) the radiation beams intensity. Optimization techniques are used for customizing the radiation dose in order to maximize the dose to the tumor while simultaneously minimizing the dose for the surrounding healthy tissues. As these goals are conflicting by their nature, IMRT treatment planning belongs to the class of multiobjective optimization problems.

5.1.1 Multiobjective Model for IMRT Beam Intensity Optimization

Various optimization problems, arising in IMRT design and treatment planning, are reviewed in [88, 43, 124] and can be classified [43] to:

- geometry problem – selection of beam angles;
- intensity problem – computation of an intensity map for each selected beam angle;
- realization problem – optimizing a sequence of configurations of the multileaf collimator.

We discuss only the intensity problem in this section assuming that the beam angles are already fixed. This problem is also called *beam intensity optimization* or *optimization of intensity maps.*

In addition to the tumor, surrounding healthy organs are also affected by the treatment. Patient anatomy is represented via a collection of voxels V_i, those are small 3D cubes. A voxel is a 3D analog of a 2D pixel and medical resolution for voxels is usually around 3 mm. Treatment beam is decomposed into a grid of "beamlets" that have different intensities.

For calculating the dose, we divide the patient's body into n voxels and the beams are discretized into m beamlets. Then

$$d = Dx,$$

where $d \in \mathbb{R}^n$ is a dose vector with components d_i corresponding to the dose deposited in voxel i; $x \in \mathbb{R}^m$ is a vector of beamlet intensities with each component x_j representing the intensity of beamlet j; $D \in \mathbb{R}_+^{n \times m}$ is the dose deposition matrix with the elements D_{ij} representing the dose deposited in voxel i due to unit intensity in beamlet j. It is assumed that D is given. Matrix D can be partitioned and reordered into sub-matrices $D_k \in \mathbb{R}_+^{n_k \times m}$ ($\sum_{k=1}^{N} n_k = n$) according to the rows corresponding to organs (or structures) that are indexed by k. It is usually convenient to have $k = 1$ for the tumor and $k = 2, \ldots, N$ for each organ at risk.

Constraints in the IMRT beam intensity optimization problem are defined [31] as:

$$\begin{array}{rcll} d & = & Dx, & \\ d_i & \leq & u_k, & \forall\, i \in V_k,\ k = 1, \ldots, N, \\ d_i & \geq & l_k, & \forall\, i \in V_k,\ k = 1, \ldots, N, \\ x & \geq & 0, & \end{array} \tag{5.1.1}$$

where d is a vector of doses d_i in each voxel i, D is the dose deposition matrix, and x is the vector of beamlet intensities to be determined. Bounds u_k and l_k are the maximum and minimum voxel dose values for organ k and those are hard constraints. Physicians specify prescribed doses for the tumor, each organ at risk and the normal tissue that are used to construct lower and upper bounds on the dose to tumor voxels as well as upper bounds (as $l_2, \ldots, l_N = 0$) on the dose to organs at risk and normal tissue voxels.

Defining an appropriate objective function for IMRT treatment planning optimization is a challenging task as it depends on many dosimetric and biological factors [109]. *Dose criteria* are based on dose distributions and involve objectives or constrains on the dose delivered to voxels of the tumor and normal tissues, see, e.g., equations (5.1.2) and (5.1.3). In many cases pure dose-based criteria are not sufficient. The observation that an organ response to radiation is a function of dose and the volume receiving the dose, led to the *dose-volume criteria* that are based on dose-volume histograms (DVH). Intuitively, the dose-volume criteria include limits on the fractions of organs that are allowed to receive certain

dose. Dose and volume criteria are often referred to as *physical criteria* as those are based on measurable physical quantities [124]. It was noted that physical criteria may be too restrictive and not properly describe the biological response of organs to irradiation patterns. Consequently, the dose-volume criteria were supplemented with biological (or dose-response based) criteria [109]. One of the biological indices is the concept of *equivalent uniform dose* (EUD), see, e.g., [109] for an overview on the subject. Other *biological criteria* are tumor control probability (TCP) and normal tissue complication probability (NTCP) [125].

The quality of dose distribution in an organ is measured by an evaluation function f. This function depends on the dose distribution and the assumptions we make about the organs at risk and about the tumor. The evaluation function f can be expressed as the deviation from the specified threshold. As the result, the most common choices for *physical criteria* (convex voxel-based criteria [125]) are

$$f_k(d) = \frac{1}{|V_k|} \sum_{i \in V_k} |d_i - \delta_k|^{q_k}, \quad q_k \in [1, \infty), \tag{5.1.2}$$

for tumors, and

$$f_k(d) = \frac{1}{|V_k|} \sum_{i \in V_k} \max\{d_i - \delta_k, 0\}^{q_k}, \quad q_k \in [1, \infty), \tag{5.1.3}$$

for critical structures (organs at risk). Here, d_i is the dose to voxel i, $|V_k|$ is the number of voxels in the structure k, δ_k is a dose threshold, q_k is a shape parameter (usually $q_k = 1, 2$, but it can also have higher values) corresponding to each voxel in the structure k.

Equations (5.1.2)-(5.1.3) are the examples of physical criteria based on convex penalty functions as those penalize the dose deviations delivered to a structure [124]. Essentially, function f in (5.1.2)-(5.1.3) penalizes dose deviations from the prescribed threshold value by the means of different norms (depending on the value of q_k). For instance, the 1-norm and the 2-norm lead to linear and quadratic objective functions, respectively. Formulations (5.1.2)-(5.1.3) allow measuring and optimizing homogeneity of the target dose distribu-

tion represented by mean-absolute deviation when $q_k = 1$, or standard deviation when $q_k = 2$.

Another approach for quantifying the quality of the dose distribution in an organ is the biological impact. Among the measures of biological impact is the EUD, thus we are going to use it as our measure. The EUD is defined as the uniform dose that would lead to the same biological effect as the given non-uniform dose in a particular organ. A number of objective functions f based on EUD-type measures are described in [135] and we are going to briefly discuss those below.

Generalized EUD (gEUD) was introduced by Niemierko [108] and we refer to it as just EUD hereafter. The objective function f for the structure (organ) k is denoted by EUD^k. The heterogenous dose distribution for an organ is translated into a single value by computing EUD^k as

$$\text{EUD}^k_{p_k} = \left(\frac{1}{|V_k|} \sum_{i \in V_k} d_i^{p_k} \right)^{1/p_k}, \tag{5.1.4}$$

where $|V_k|$ is the number of voxels in an organ k, d_i is the dose delivered to voxel i in that structure, and p_k is an organ-specific number. In fact, $p_k = 1$ refers to the parallel organ characteristic, while larger values of p_k ($p_k \to \infty$) point to the serial structure of that organ [108]. In mathematical terminology, $p_k = 1$ is translated to 1-norm in the former case, as opposed to ∞-norm in the later case.

In [25] convexity properties of the EUD function were studied. It was shown that generalized EUD function EUD_p is convex or concave depending on its only parameter p. When $p \geq 1$, minimizing EUD_p on a convex feasibility space leads to convex optimization problem. Similarly, when $p < 1$, maximizing EUD_p on a convex feasibility space leads to convex optimization problem. Moreover, for $p = 1$ linear function is obtained.

We are going to discuss the EUD approach in more detail below. However, we first turn our attention to the multiobjective nature of IMRT beam intensity optimization problem. The general goal of multiobjective optimization in IMRT treatment planning is determining the doses of beams for the voxels of

each organ, such that the delivered dose has maximum effect on the cancerous organ, while minimizing the damage to healthy tissues. References [84, 124] and [70] describe the multiobjective nature of treatment plan optimization problems, and [84] contains the most detailed overview of multicriteria optimization problems appearing in IMRT. Multiple objectives appear in problem statement due to existence of a number of treatment objectives, as well as due to multiple body organs affected by the treatment.

Many authors formulated the beam intensity optimization problem as a multiobjective LO problem. To solve this type of problems, they try to approximate the whole Pareto front or it parts [32, 131]. For instance, in [32] the authors formulate a multiobjective linear IMRT problem with three objectives and compute an approximation of the Pareto efficient surface. We take another approach in this section and compute the subset of Pareto front exactly using parametric optimization techniques described in Chapter 2 and employing the links between multiobjective and parametric optimization established in Chapter 3 with implementation ideas from Chapter 4.

The case study in this section demonstrates the use of parametric optimization techniques for multiobjective linear IMRT treatment planning problems. The goal is to show that parametric optimization can be used instead of approximation techniques to compute exact Pareto fronts. Due to the high dimensionality of the optimization problem, practically it may be worthwhile to implement approximation techniques similar to [32, 71, 131] to obtain an approximate Pareto front and employ parametric optimization to compute the structure of the exact Pareto front only for the areas of interest that are clinically relevant.

A general example of EUD-type multiobjective optimization problem for IMRT treatment planning is formulated in [125] as:

$$\min_{x \geq 0} \{-\mathrm{EUD}^1_{p_1}(x), \ldots, -\mathrm{EUD}^T_{p_T}(x), \mathrm{EUD}^{T+1}_{p_{T+1}}(x), \ldots, \mathrm{EUD}^N_{p_N}(x)\},$$

where $k = 1, \ldots, T$ indexes target (tumor) structures and $k = T+1, \ldots, N$ indexes critical structures. As we see from the problem constrains (5.1.1), the dose vector d can be eliminated from the formulation and the problem can be expressed solely by using the variables x.

Based on (5.1.4), in [139] an alternative formulation of the EUD-type objective or constraints is introduced, and later on used in a number of publications, including [31]. This alternative expression for EUD is called "linear EUD" and is denoted as αEUD. In that model, the objective function for a healthy organ k is denoted by αEUD^k and is defined as

$$\alpha EUD^k = \alpha_k \cdot \max_{i \in V_k} d_i + (1 - \alpha_k) \cdot \frac{1}{|V_k|} \sum_{i \in V_k} d_i, \tag{5.1.5}$$

where $|V_k|$ is a number of voxels in structure k, d_i is the dose delivered to voxel i in that organ, and $\alpha_k \in [0, 1]$ is a region-specific parameter that depends on the critical structure. Even though the max term in this definition appears to be nonlinear on the first glance, it can be linearized with the help of auxiliary variables and additional constraints, so, model (5.1.5) is linear. Note that for the tumor, the max term in (5.1.5) is replaced by the min term [135].

The αEUD approach is called "max and mean model" as the formulation (5.1.5) can be rewritten as $\alpha\text{EUD}^k = \alpha_k \cdot \text{EUD}^k_\infty + (1 - \alpha_k) \cdot \text{EUD}^k_1 = \alpha_k \cdot d^{\max} + (1 - \alpha_k) \cdot d^{\text{mean}}$, where $d^{\max}$ is the maximum and d^{mean} is the mean radiation dose delivered to a healthy organ. Parameter α_k is organ specific and is determined from Emami tables [46]. In solving practical problems, value of α_k is predefined by the radiotherapist and may defer from one case to another.

Let us consider a prototype multiobjective LO problem with two objective functions of the αEUD-type and the associated constraints (5.1.1). Two objective functions correspond to maximizing αEUD^1 for the tumor ($k = 1$) and minimizing αEUD^2 for one organ at risk ($k = 2$). The problem formulation is as follows:

$$\begin{aligned}
\min_{x,d} \quad & \left\{ -\left(\alpha_1 \min_{i \in V_1} d_i + (1 - \alpha_1) \frac{1}{|V_1|} \sum_{i \in V_1} d_i \right), \left(\alpha_2 \max_{i \in V_2} d_i + (1 - \alpha_2) \frac{1}{|V_2|} \sum_{i \in V_2} d_i \right) \right\} \\
\text{s.t.} \quad & d = Dx \\
& d_i \leq u_k, \quad \forall i \in V_k,\ k = 1, 2 \\
& d_i \geq l_k, \quad \forall i \in V_k,\ k = 1, 2 \\
& x \geq 0.
\end{aligned} \tag{5.1.6}$$

The objective function αEUD^k may have an upper bound U_k adding another restriction on the Pareto front.

We are going to use problem (5.1.6) as the prototype multiobjective optimization problem for the case study in Section 5.1.2.

5.1.2 Parametric IMRT Case Study

For illustrative purposes we consider three body organs, namely a *bladder*, *prostate* and *rectum*. Prostate is affected by cancer, while two other structures are healthy. We are interested in finding the beams radiation dose to be delivered to the three organs with the goal that it has maximum effect on the tumor in a prostate (associated αEUD needs to be maximized) and sparing the two other structures (associated αEUDs are minimized). We use the test data produced in CERR computational environment for radiotherapy research [38]. CERR was used for pre-selecting the beam angles and calculating the dose deposition matrix D.

In this case study, six sources of equispaced beams are considered. Organs are indexed by k with the convention that $k = 1$ corresponds to the prostate (tumor), $k = 2$ – to the rectum and $k = 3$ – to the bladder. The matrices D_k are of high dimension (D_1 is 3882×165, D_2 is 10033×165 and D_3 is 15001×165) and all of those are sparse matrices. Correspondingly, the number of voxels for the organs is $|V_1| = 3882$ (prostate), $|V_2| = 10033$ (rectum) and $|V_3| = 15001$ (bladder).

The multiobjective LO problem formulation for the case study is as follows:

$$\begin{array}{ll} \min\limits_{x,d} & -\left(\alpha_1 \min_{i\in V_1} d_i + (1-\alpha_1)\frac{1}{|V_1|}\sum_{i\in V_1} d_i\right) \\ \text{s.t.} & \alpha_2 \max_{i\in V_2} d_i + (1-\alpha_2)\frac{1}{|V_2|}\sum_{i\in V_2} d_i \le \varepsilon_R \\ & \alpha_3 \max_{i\in V_3} d_i + (1-\alpha_3)\frac{1}{|V_3|}\sum_{i\in V_3} d_i \le \varepsilon_B \\ & d = Dx \\ & d_i \le u_k, \quad \forall\, i \in V_k,\ k = 1,2,3 \\ & d_i \ge l_k, \quad \forall\, i \in V_k,\ k = 1,2,3 \\ & x \ge 0. \end{array} \tag{5.1.7}$$

Multiobjective optimization problem (5.1.7) is written in the form of ε-constrained formulation with the parameter ε_R corresponding to the upper bound

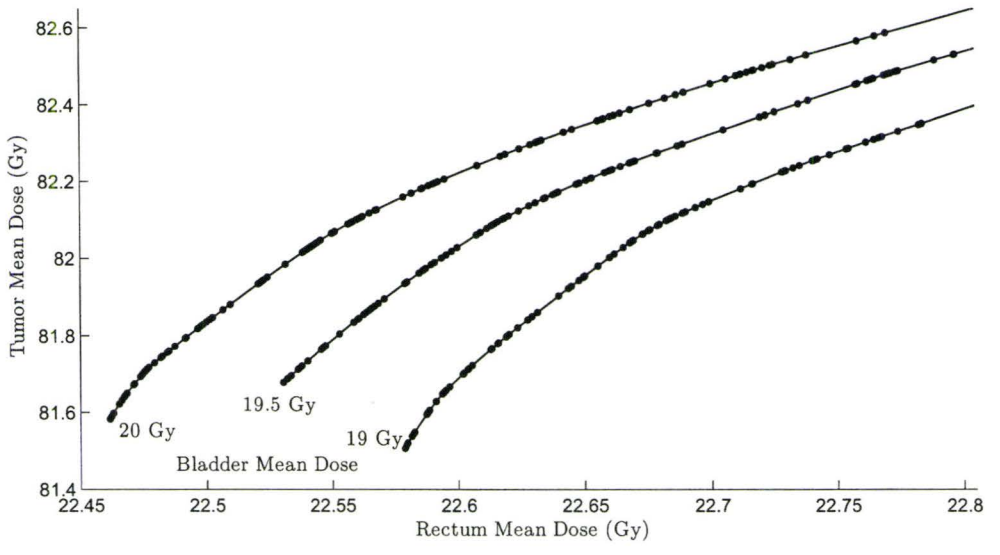


Figure 5.1: 2D Pareto Frontier for Prostate Cancer IMRT Treatment Planning Considering the Trade-Off between Tumor Mean Dose and Rectum Mean Dose.

on the αEUD of the rectum and ε_B – of the bladder. For our computational results, we assigned the following values $l_1 = 75\,\text{Gy}$, $l_2 = l_3 = 0$, $u_1 = 90\,\text{Gy}$, $u_2 = 76\,\text{Gy}$ and $u_3 = 78\,\text{Gy}$. We also have taken $\alpha_2 = \alpha_3 = 0$ in the formulation (5.1.7). For our experiments shown in Figures 5.1-5.3 we have assigned $\alpha_1 = 0$, while for the experiment in Figure 5.4 $\alpha_1 = 1$. Parametric LO problem (5.1.7) was reformulated in the standard form and solved using parametric optimization techniques described in Chapter 2.

Figure 5.1 shows Pareto frontiers for αEUDs considering the trade-off between tumor mean dose and rectum mean dose, when the bladder mean dose is fixed at the levels of 19 Gy, 19.5 Gy and 20 Gy. Figure 5.1 depicts a family of 2D Pareto fronts, where the solid dots correspond to transition points. The tumor mean dose depends on the rectum mean dose linearly on the invariancy intervals between transition points.

Figure 5.2 shows the approximation of the 3D Pareto front for αEUDs, considering the trade-off between tumor mean dose, rectum mean dose, and bladder

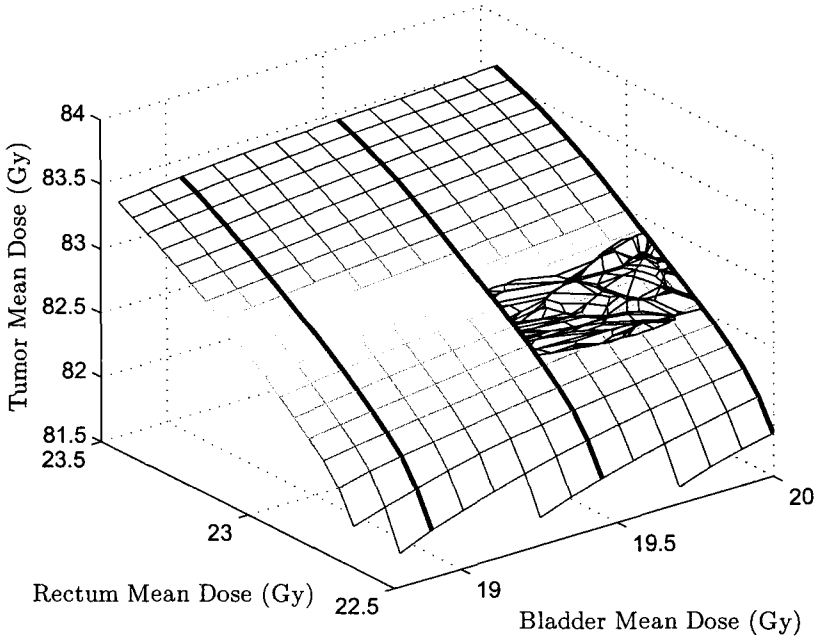


Figure 5.2: The 3D Pareto Surface for Prostate Cancer IMRT Treatment Planning.

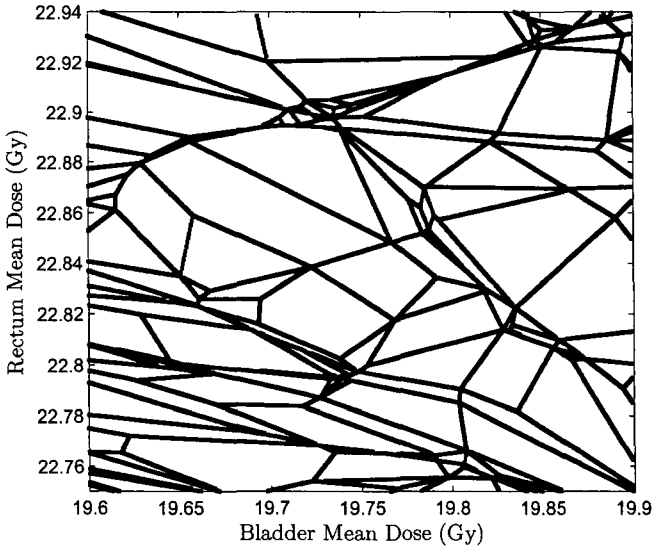


Figure 5.3: Invariancy Regions Corresponding to the 3D Pareto Front for Prostate Cancer IMRT Treatment Planning.

mean dose. The solid dark curves in Figure 5.2 are the exact 2D Pareto fronts depicted in Figure 5.1. Using parametric optimization we have also computed the exact 3D Pareto front for the region of interest. That Pareto front is drawn in Figure 5.2 on the top of the approximated one. Invariancy regions corresponding to the Pareto front in Figure 5.2 are shown in Figure 5.3.

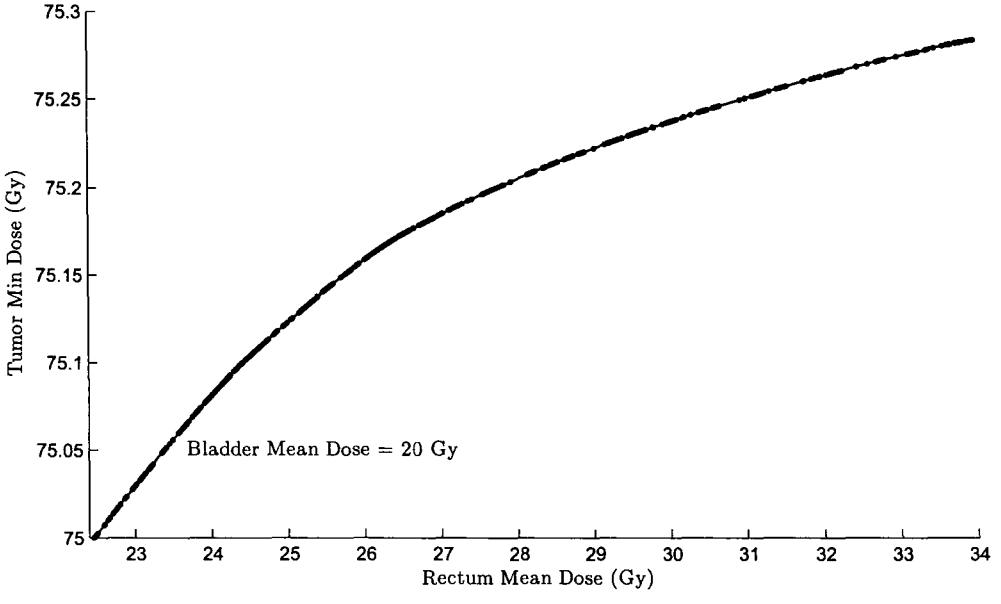


Figure 5.4: 2D Pareto Frontier for Prostate Cancer IMRT Treatment Planning Considering the Trade-Off between Tumor Min Dose and Rectum Mean Dose.

Finally, Figure 5.4 shows the Pareto frontier for αEUDs, considering the trade-off between tumor min dose and rectum mean dose, when the bladder mean dose is fixed at the level of 20 Gy.

5.1.3 Parametric Model Extensions for IMRT Beam Intensity Optimization

While we have considered a number of multiobjective linear optimization problems for IMRT treatment planning in this section, the class of optimization prob-

lems used in IMRT design is much wider. Moreover, many of the formulations are multiobjective non-linear optimization problems. We describe below a couple of extensions that can be the potential applications of parametric optimization.

As we have seen in Section 5.1.1, many models are based on minimizing the deviation from the prescribed dose for the tumor. For instance, an objective function for the tumor dose in these formulations can be expressed by equation (5.1.2). If the squared deviation from the prescribed tumor dose is minimized and all other problem objectives and constraints are linear, e.g., linear EUD-type objectives and constraints, linear bounds, etc. are used, this formulation results in a multiobjective QO problem. Solution techniques for such parametric QO problems are developed in Chapter 2.

As all treatment planning problems involve uncertainty in tumor position due to organ motion and patient displacement, robust optimization techniques can be applied to offset the adverse effects of those uncertainties. Robust optimization formulations for IMRT treatment planning are described in [27], where the authors formulate robust counterparts of linear optimization problems for treatment planning. Those robust formulations are SOCO problems as ellipsoidal uncertainty in tumor position was assumed. Multiobjective optimization can be done on the top of the robust formulations of [27]. In that case we obtain parametric SOCO problem. A variant of SOCO framework for IMRT treatment planning is also presented in [159], where the optimization problem includes generalized moment constraints as generalized EUD-type constraints. Parametric SOCO techniques can be applied to those formulations as well. Ideas and preliminary results about solving parametric SOCO problems are discussed in Chapter 8.

5.2 Optimization of Multi-Drug Composition for the Most Efficacious Action

Chinese herbal medicine has been used in East Asia for twenty centuries and during the past decades it is becoming more and more accepted by health care

professionals and patients in North America and Europe [146, 147]. Compared to western drug therapy, often administered in the form of a single active chemical ingredient, Chinese herbal medicine typically combines several herbs containing up to hundreds of chemical compounds each, and having complex relationships between their bioactivities. This property of Chinese herbal medicine makes it difficult to quantify the dose-response relation in herbal therapy, or even to isolate the active ingredients responsible for the therapeutic mechanism.

5.2.1 Composition-Activity Modeling and Subset Selection

Modeling *quantitative composition-activity relationships* (QCAR) is a central goal in the mathematical analysis of existing drugs and for the design of new biologically active compounds. In the area of herbal therapy, references such as [26, 146, 147, 148, 154] use QCAR to predict Chinese herbal medicine's bioactivity from its composition based on computational analysis. Multiple linear regression is the most widely used method to deduce a linear QCAR, i.e., a linear mapping from the composition vector of the drug or herbal mixture to the biological activity or response. However, it has been suggested that potentially important synergies between active components in herbal remedies are missed by this approach, and this may render the models unreliable, or at least limited in scope.

Moving away from the limitations of linear models, there seems to be a general agreement that the response variable y should be a function of dose d, and $z = (z_1, z_2, \ldots, z_m)$, the component fractions z_j of all assumed active components in the medication. The dose $0 < d < \infty$ can be measured in units of milligrams per milliliter – the strength or potency of the medication. So, for example, $z_j \geq 0$, $j = 1, \ldots, m$, and $\sum_j z_j = 1$ may be assumed, and the z_j are unit-free parameters. Further, the effect of dose d should satisfy the Michaelis-Menten chemical kinetics equation relating concentration to rate of the chemical activity [14]. It is reasonable to assume that the concentration is a combination

of dose d and component fractions $q(z)$, where $q(z)$ is a general quadratic expression in z to allow for individual component activity and pairwise synergistic or inhibitory effects in the mixture. For the ease of explanation and to reduce computational complexity, we assume that the observed response y is modeled as $y = d \cdot q(z)$. For convenience, we ran our numerical experiments by fitting response y at dose $d = 1.0$ rather than the full Michaelis-Menten model.

Our task is to determine an expression for $q(z)$, the unknown quadratic expression

$$q(z) = \alpha_0 + \sum_{j=1}^{m} \alpha_j z_j + \sum_{j=1}^{m} \sum_{k=1}^{m} \beta_{jk} z_j z_k. \tag{5.2.1}$$

The general model described by equation (5.2.1) can be used for answering a number of questions:

- Identify the variables z_j and their combinations $z_j z_k$ that are the most active, or most relevant for the observed response in a given herbal mixture. It is generally assumed that the number of active components may be as few as 4 or 5 for most of the herbal mixtures of interest.
- Fit the response surface using multiple linear regression with respect to the most active components, and analyze the resulting model for sensitivity to measurement error on these critical components. Such issues are naturally of interest for the production, sale and regulation of Chinese herbal medicines, especially in the West.
- Determine the mixture of the components that leads to the maximal response given the model for a specified dose d.

To fit the model (5.2.1) to the data on the components $z = (z_1, z_2, \ldots, z_m)$ we denote the unknown model coefficients by $x = (\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_m, \beta_{1,2}, \beta_{1,3}, \ldots, \beta_{1,m}, \beta_{2,3}, \ldots, \beta_{2,m}, \ldots, \beta_{m-1,m})^T$ and known data on the component fractions as

$$a = (1, z_1, z_2 \ldots, z_m, z_1 z_2, z_1 z_3, \ldots, z_1 z_m, z_2 z_3, \ldots, z_2 z_m, \ldots, z_{m-1} z_m)^T. \tag{5.2.2}$$

Now, a goodness of fit criterion $\|y_i - q(z^i)\|$, $i = 1, \ldots, N$ can be minimized as $\min_x \|y_i - a_i^T x\|$ to compute the vector of unknown coefficients x.

The sample data set is provided by the SinoVeda company and is described in [14]. It contains $N = 25$ response samples and $m = 15$ components in the mixture. Response y was calculated from $y = d \cdot q(z)$ for a known dose $d = 1$. For the data set, the number of regressors is equal to $n = 1 + m + (m^2 - m)/2 = 1 + 15 + (15^2 - 15)/2 = 121$.

It was shown in [14] that principal component analysis may not always work to reduce dimensions in underdetermined systems. So, our approach is to build the regression model for the problem with *subset selection*. The goal here is to find the best models which are built on subsets of a full multiple regression model. This is known as "attribute selection", "feature selection" or "variable selection" in the literature.

In our case, we have too many predictors or variables (with many possible interactions) making it difficult to find a good model. In general, this is an unsolved problem in statistics: there are no magic procedures to get the best model. To implement the subset selection we need a criterion or benchmark to compare two models and a search strategy. With a limited number of predictors, it is possible to search all possible models by using exhaustive search.

The problem to be solved is as follows. For a given subset $S \subseteq 1, 2, \ldots 15$ representing a choice of reduced predictors that are the variables in our model, let z_S^i denote the data vector z^i with only the S variables present. Recall that we have $i = 1, 2, \ldots, N = 25$ data vectors. Having chosen S, we proceed with the regression

$$\hat{q}_S := \operatorname{argmin} \|y_i - q(z_S^i)\|_2^2 := \operatorname{argmin} R^2, \tag{5.2.3}$$

where R^2 is the coefficient of determination that measures the goodness of fit of the model, and it serves as one of possible criteria for subset selection.

The review of different criteria for the subset selection problem can be found in [53]. Possible criteria to compare different subsets S are:

- R^2: may not be the best criterion as it always increases with the model size and, as a result, the "optimum" is getting to take the biggest model;
- Adjusted R^2: it is a modification of R^2 that adjusts for the number of

predictor variables in the model. Adjusted R^2 is better than R^2 as it "penalizes" bigger models. Adjusted R^2 increases only if a new variable improves the model more than it would be expected by chance;

- Mallows' C_p: Mallows [93] developed a method to select adequate models by plotting a special statistic against the number of variables + 1. Mallows' C_p statistic is used in multiple regression analysis to select models that contain smaller numbers of predictors from a larger number that is available for inclusion. Small values of C_p indicate better models;
- Akaike's Information Criterion (AIC) and Schwarz Bayesian Information Criterion (BIC). AIC and BIC are measures of the goodness of fit of an estimated statistical model. $AIC = -2 \cdot \ln L + 2k$ and $BIC = -2 \cdot \ln L + k \cdot \ln N$, where L is the maximized value of the likelihood function for the estimated model, N is the sample size and k is the number of regressors including the constant.

Possible search strategies include:

- 'Best subset': search all possible models and take the one with highest adjusted R^2 or lowest C_p;
- 'Stepwise' (forward, backward or both): choose an initial model and take the biggest jump up or down in the selected criterion.

For instance, a function in the R software package for statistical computing [48] is available, that performs an exhaustive search for the best subsets of the variables in z for predicting q in linear regression (5.2.3) using an efficient branch-and-bound algorithm based on [104]. The 'best subset' selection as implemented in R can use either adjusted R^2, C_p or BIC and does exhaustive search under the branch-and-bound algorithm. Exhaustive search is equivalent to integer programming, which is expensive and only works for small model sizes.

We ran the best subset selection with adjusted R^2 as our criterion. Figure 5.5 shows the plots of the optimal subsets for dose=1.0. The regressors (variables) are on the horizontal axis and adjusted R^2 values are shown on the vertical axis. The plot includes four best subsets that were produced by exhaus-

tive search of each size. The subset sizes of up to 5 regressors are shown (up to 6 counting the intercept). The bars on Figure 5.5 correspond to the variables included in the subset and those variables are ranked based on the adjusted R^2 criterion.

The crucial decision to make here is how many variables to select. This decision is usually based on the knowledge of the problem in hand, drug-design experience and additional information available. For instance, if we use the SinoVeda data set and believe that only five variables (plus the intercept) should be selected, we get the following selected variables that represent components in the mixture:

$$z_2,\ z_{10},\ z_{12},\ z_9z_{13},\ z_{10}z_{12}.$$

Inclusion of these variables into the regression produces the maximal adjusted R^2 as shown at the very top of Figure 5.5. Please note that if z_9z_{13} is selected as one of the regressors for the model, than the interaction between components 9 and 13 in the mixture is important for explaining the activity of the drug. The resulting model in our case is:

$$y = \alpha_0 + \alpha_2 z_2 + \alpha_{10} z_{10} + \alpha_{12} z_{12} + \beta_{9,13} z_9 x_{13} + \beta_{10,12} z_{10} z_{12}.$$

In the remainder of Section 5.2.1 we show the results obtained by running the exhaustive search for subset selection. A fitted model with subset of size five for our data set is:

$$y = 0.33 + 0.19z_2 + 0.12z_{10} + 0.24z_{12} + 0.09z_9z_{13} - 0.23z_{10}z_{12}.$$

Note that interaction between components 10 and 12 in the mixture exhibits inhibition effect while interaction between components 9 and 13 exhibits synergism. The intercept and the coefficients for components 2, 10 and 12 appear with positive sign indicating that the corresponding components increase the activity of the drug. In general case, some components may suppress the activity.

Figure 5.6 shows the response surfaces for different pairs of components. Concentrations of other components in the mixture are kept constant. In that

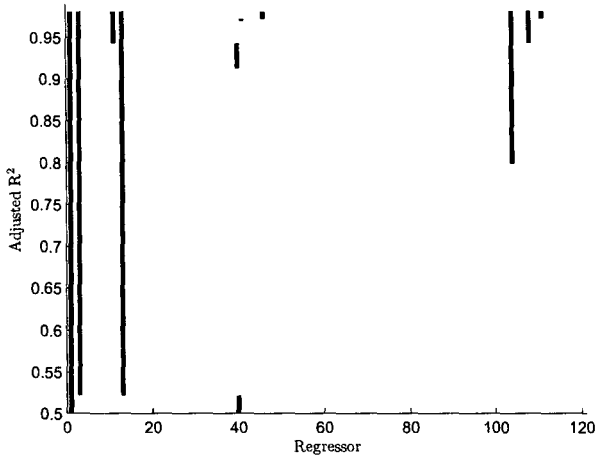


Figure 5.5: Plots of Optimal Subsets for Dose=1.

case, we can analyze the pair-wise interactions between components to find their influences on the drug activity.

We would like to stress that the exhaustive search is not the best algorithm to solve subset selection problems and we included it here for illustration and comparison purposes. For the illustrative data set used in this thesis we have $N = 25$ samples and $n = 121$ regressors, but for the practical data sets the number of samples can be around $N = 500$ and the number of components in the samples can be as high as 500 resulting in $n = 1+500+(500^2-500)/2 = 125,251$. In Section 5.2.2 we illustrate how optimization techniques and, in particular, parametric optimization can be used to solve subset selection problems for drug-design.

5.2.2 Regularized Regression via Parametric Optimization

It turns out that we can combine subset selection for model parameters, as described in Section 5.2.1, with linear regression curve fitting. The technique for that is called *least absolute shrinkage and selection operator* (LASSO) regression

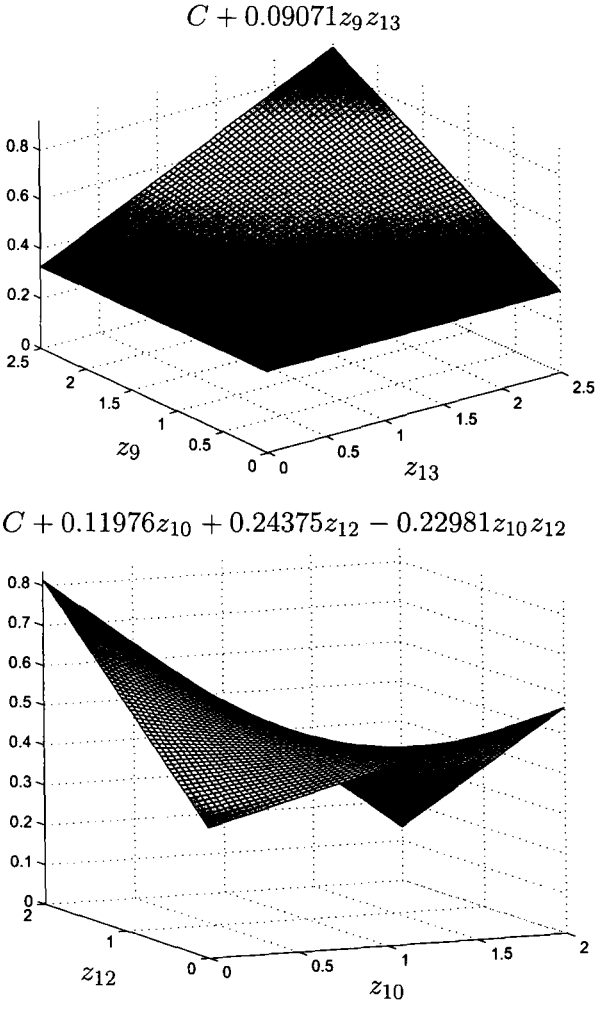


Figure 5.6: Response Surfaces for Pairs of Components.

model [140].

As we have an underdetermined regression model where the number of variables $n = \frac{m(m+1)}{2} + m + 1$ exceeds the number of samples N, we need to find a model having small regression error with a number of variables much less than n. The more sparse model we get, the more likely is that the predictors z have causal relationship to the dependant variable y [118]. The LASSO uses ℓ_1 norm as the shrinkage function as it turns out that the ℓ_1 regularization term in the

objective function produces sparse solutions. We discuss regularized optimization in more detail in Section 6.2, here we keep the explanations to minimum. So, the convex optimization problem that minimizes the Euclidian norm of residuals combined with ℓ_1-norm regularization produces sparse solutions for the linear regression model:

$$\min_x \|y - Ax\|_2^2 + \lambda\|x\|_1. \tag{5.2.4}$$

Here, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^N$, A is an $N \times n$ matrix with the rows a as defined in (5.2.2), and λ is a non-negative tuning parameter. In addition, $\|v\|_2$ is the Euclidian norm of v and $\|v\|_1 = \sum_i |v_i|$ is the ℓ_1-norm of v.

The choice of ℓ_1-norm is a heuristic. A more natural choice would be ℓ_p for some $p < 1$ since minimizing $\|x\|_p$, subject to holding $\|x\|_2$ strictly away from zero will favor sparsely supported vectors. However, as ℓ_p norms are not convex for $p < 1$, we would be left with a non-convex objective and obvious difficulties with the above optimization. Consequently, ℓ_1 norms are a compromise choice for regularization. Further discussion on this point can be found in Section 6.2 and in [15] pp. 308–310.

By varying the parameter λ we can control the trade-off between $\|y - Ax\|_2$ and the number of non-zero elements (sparsity) of the vector x controlled by the term $\|x\|_1$. The optimization problem (5.2.4) is the parametric optimization problem with parameter λ. Selection of the value for the parameter λ can be based on cross-validation [118] or the historical experiences of drug design. However, instead of fixing λ to a particular value, we would like to compute a trade-off surface between the goodness of fit and sparsity.

Problem (5.2.4) can be solved by the following multiobjective optimization problems [50]:

$$\begin{array}{ll} \min_x & \|x\|_1 \\ \text{s.t.} & \|y - Ax\|_2^2 \leq \epsilon, \end{array} \tag{5.2.5}$$

and

$$\begin{array}{ll} \min_x & \|y - Ax\|_2^2 \\ \text{s.t.} & \|x\|_1 \leq \varepsilon, \end{array} \tag{5.2.6}$$

where ϵ and ε are non-negative parameters. Problem (5.2.5) is parametric SOCO

problem, while problem (5.2.6) is parametric QO problem. Paper [50] reviews algorithms for solving different reformulations of problem (5.2.4) with fixed parameter values and proposes a new gradient projection algorithm for that purpose. Formulation (5.2.6) is called LASSO, while formulation (5.2.4) is also known as basis pursuit denoising problem [15].

Similarly to [50], we reformulate the problem as parametric QO problem by splitting the variable x into positive and negative parts:

$$x = u - v,\ u \geq 0,\ v \geq 0.$$

Now, $\|x\|_1 = \mathbf{1}^T u + \mathbf{1}^T v$ and

$$\begin{array}{rl} \min_{u,v} & \|y - A(u-v)\|_2^2 \\ \text{s.t.} & \mathbf{1}^T u + \mathbf{1}^T v \leq \varepsilon, \\ & u \geq 0, \\ & v \geq 0, \end{array} \tag{5.2.7}$$

or

$$\begin{array}{rl} \min_{\tilde{x}} & c^T \tilde{x} + \frac{1}{2}\tilde{x}^T B \tilde{x} \\ \text{s.t.} & \mathbf{1}_{2n}^T \tilde{x} + t = \varepsilon, \\ & \tilde{x} \geq 0, \end{array} \tag{5.2.8}$$

where

$$\tilde{x} = \begin{bmatrix} u \\ v \end{bmatrix},\ b = A^T y,\ c = \begin{bmatrix} -b \\ b \end{bmatrix}, B = \begin{bmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{bmatrix}.$$

Table 5.1 shows the output of the parametric solver for the problem (5.2.8). As we have split the variables and introduced slack variable t, the total number of variables for the problem (5.2.8) is equal to $2 \cdot 121 + 1 = 243$. We print only subsets $\mathcal{B}$ and $\mathcal{T}$ of the optimal partition in Table 5.1 as the cardinality of set $\mathcal{B}$ is equal to the cardinality of the optimal solution for that invariancy interval. Note that for the invariancy interval $(1.42771, +\infty)$ we obtain weak Pareto optimality as slack variable t belongs to the set $\mathcal{B}$ of the optimal partition ($t > 0$). Optimal value function $\phi(\varepsilon)$ is quadratic on each invariancy interval and is computed in Table 5.1 and graphed in Figure 5.7. Optimal value function in Figure 5.7 is the Pareto front and it shows the trade-off between the measure of fit and the regularization term controlling solution sparsity.

Table 5.1: Parametric Solver Output for the Chinese Medicine Problem.

type	ε_ℓ	ε_u	$\mathcal{B}$	$\mathcal{T}$	$\phi(\varepsilon)$
tr. point	+0.00000	+0.00000		1	7.9684
inv. interv	+0.00000	+0.52829	1		$25.000\varepsilon^2$-28.000ε+7.9684
tr. point	+0.52829	+0.52829	1	3	0.1535
inv. interv	+0.52829	+0.57493	1 3		$7.3178\varepsilon^2$-9.3176ε+3.0336
tr. point	+0.57493	+0.57493	1 3	13	0.0955
inv. interv	+0.57493	+0.63837	1 3 13		$2.8404\varepsilon^2$-4.1692ε+1.5536
tr. point	+0.63837	+0.63837	1 3 13	10	0.0496
inv. interv	+0.63837	+0.69623	1 3 10 13		$2.1337\varepsilon^2$-3.2669ε+1.2656
tr. point	+0.69623	+0.69623	1 3 10 13	14	0.0254
inv. interv	+0.69623	+0.73149	1 3 10 13 14		$1.2683\varepsilon^2$-2.0619ε+0.8461
tr. point	+0.73149	+0.73149	1 3 10 13 14	11	0.0165
inv. interv	+0.73149	+0.75312	1 3 10 11 13 14		$1.2012\varepsilon^2$-1.9636ε+0.8102
tr. point	+0.75312	+0.75312	1 3 10 11 13 14	5	0.0126
inv. interv	+0.75312	+0.77404	1 3 5 10 11 13 14		$1.0702\varepsilon^2$-1.7664ε+0.7359
tr. point	+0.77404	+0.77404	1 3 5 10 11 13 14	61	0.0099
inv. interv	+0.77404	+0.78178	1 3 5 10 11 13 14 61		$0.7642\varepsilon^2$-1.2926ε+0.5526
tr. point	+0.78178	+0.78178	1 3 5 10 11 13 14 61	205	0.0091
inv. interv	+0.78178	+0.78607	1 3 5 10 11 13 14 61 205		$0.6188\varepsilon^2$-1.0653ε+0.4637
tr. point	+0.78607	+0.78607	1 3 10 11 13 14 61 205	5	0.0087
inv. interv	+0.78607	+0.79528	1 3 10 11 13 14 61 205		$0.6877\varepsilon^2$-1.1736ε+0.5063
tr. point	+0.79528	+0.79528	1 3 10 11 13 14 61 205	31	0.0079
inv. interv	+0.79528	+0.80624	1 3 10 11 13 14 31 61 205		$0.6609\varepsilon^2$-1.1310ε+0.4893
tr. point	+0.80624	+0.80624	1 3 10 11 13 14 31 61 205	150	0.0071
inv. interv	+0.80624	+0.80864	1 3 10 11 13 14 31 61 150 205		$0.3483\varepsilon^2$-0.6270ε+0.2861
tr. point	+0.80864	+0.80864	1 3 10 11 13 14 31 61 150 205	142	0.0069
inv. interv	+0.80864	+0.83068	1 3 10 11 13 14 31 61 142 150 205		$0.2846\varepsilon^2$-0.5240ε+0.2445
tr. point	+0.83068	+0.83068	1 3 10 13 14 31 61 142 150 205	11	0.0056
inv. interv	+0.83068	+0.83819	1 3 10 13 14 31 61 142 150 205		$0.2863\varepsilon^2$-0.5267ε+0.2456
tr. point	+0.83819	+0.83819	1 3 10 13 14 31 61 142 150	205	0.0053
inv. interv	+0.83819	+0.84215	1 3 10 13 14 31 61 142 150		$0.3273\varepsilon^2$-0.5956ε+0.2745
...					
inv. interv	+1.39370	+1.42771	1 3 10 11 13 49 ... 238		$0.0077\varepsilon^2$-0.0219ε+0.0156
tr. point	+1.42771	+1.42771	1 3 10 11 13 49 ... 238	34 45 149	0.0000
inv. interv	+1.42771	+Inf	1 2 3 4 ... 241 242 243		$0.0000\varepsilon^2$-0.0000ε+0.0000

Figures 5.8 and 5.9 demonstrate the optimal solution cardinality vs. two objective function values, namely the measure of fit $\|y - Ax\|_2$ and the regularization term $\|x\|_1$. It can be observed from the figures that the cardinality is not monotonic in the values of the regularization term.

Now, we can compare the results obtained by running exhaustive search for the subset selection to the results obtained by the parametric optimization algorithm applied to formulation (5.2.6). Figure 5.10 shows comparison of the regularized heuristic approach via parametric optimization with enumeration. Horizontal axes show the cardinality of the optimal solutions obtained by each

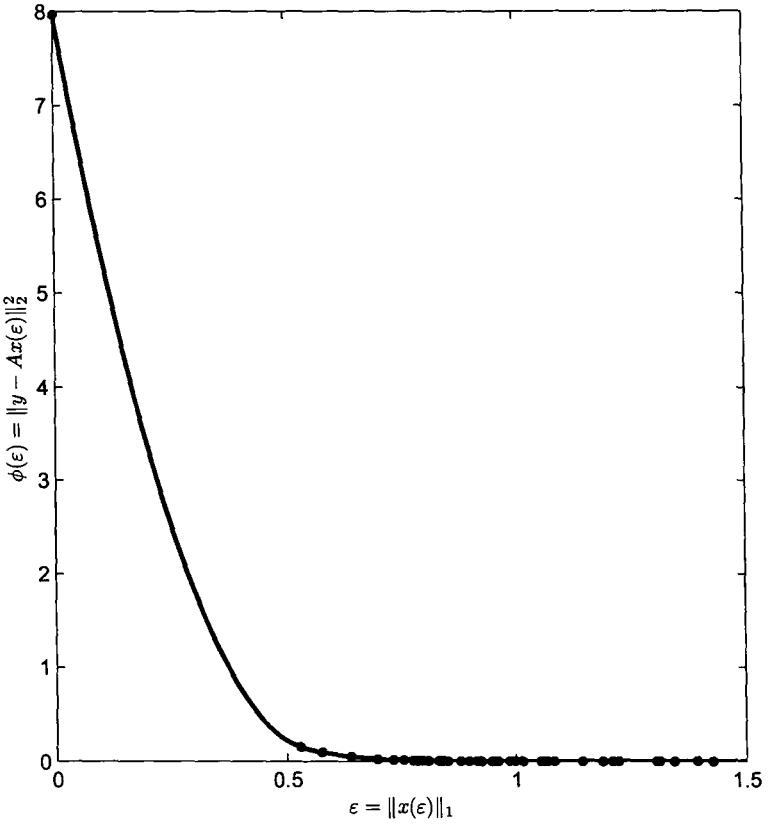


Figure 5.7: Optimal Value Function for the Chinese Medicine Problem.

of the two approaches, while the vertical axes depict the goodness of fit measure. We have computed globally optimal values by exhaustive search only for up to 8 regressors due to prohibitively long running time of the enumeration. Results in Figure 5.10 demonstrate that comparable quality solutions can be computed with parametric optimization in a fraction of time used for enumeration.

For an appropriate value of ε that corresponds to selecting 5 variables for the regression, we obtained the following model:

$$y = 0.31 + 0.19z_2 + 0.07z_9 + 0.02z_{10} + 0.16z_{12} + 0.07z_{13}.$$

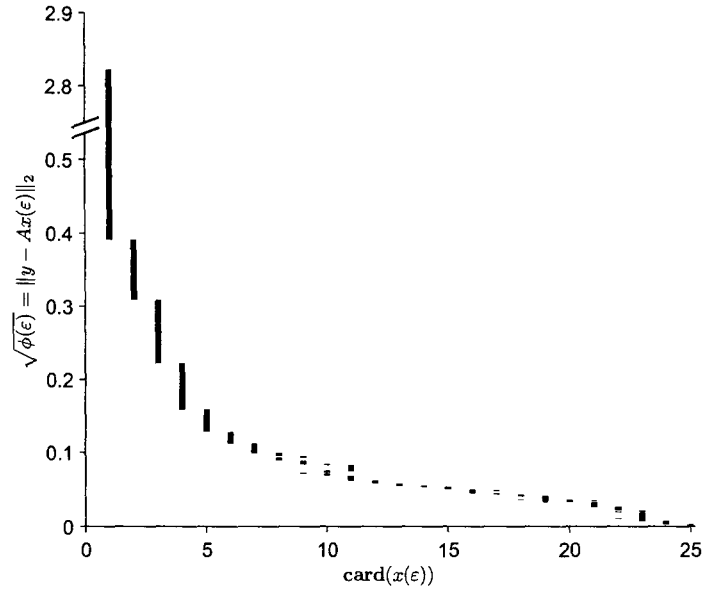


Figure 5.8: Optimal Solution Measure of Fit vs. Sparsity for the Chinese Medicine Problem.

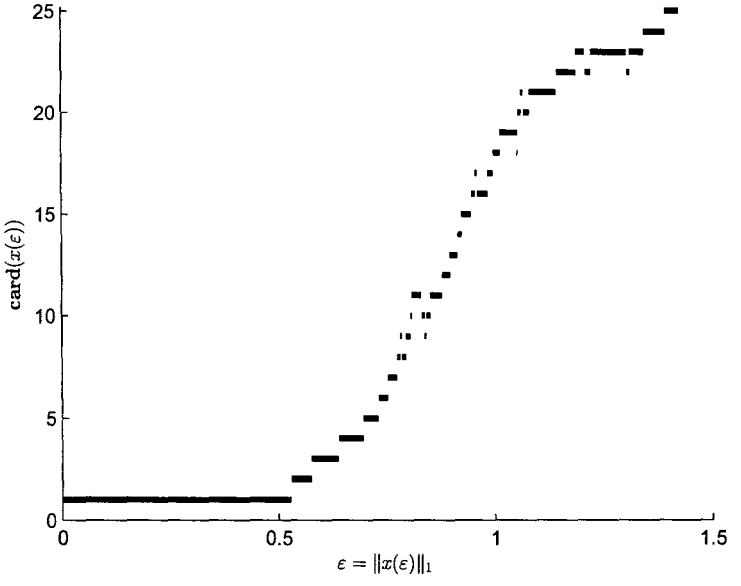


Figure 5.9: Optimal Solution Cardinality vs. Norm Minimization for the Chinese Medicine Problem.

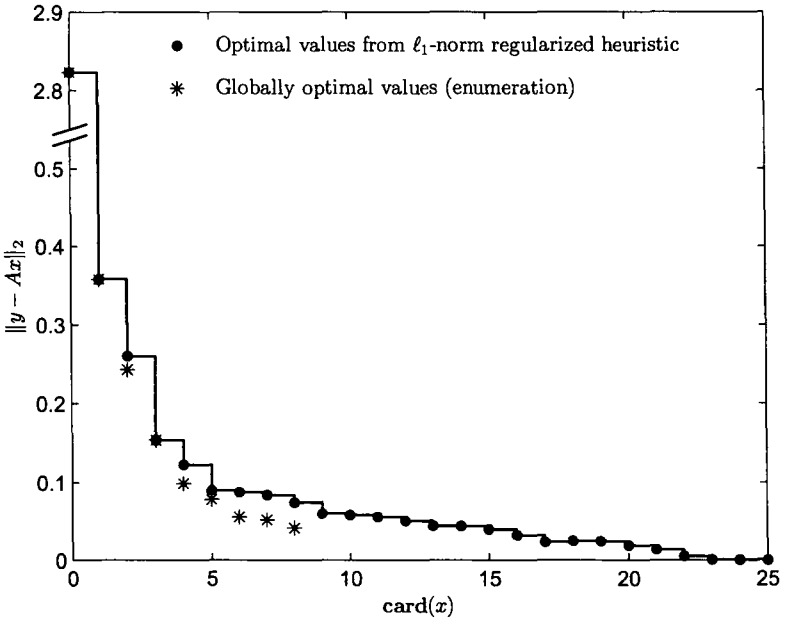


Figure 5.10: Globally Optimal and ℓ_1-Norm Heuristic Solution for the Chinese Medicine Problem.

The value of adjusted R^2 drops from 0.97 to 0.92 for this model as compared to the exhaustive search one and the coefficient for variable z_{10} is not significant. In spite of that, optimization proves to be very useful for solving large-scale subset selection problems.

Part II

Financial Models and Parametric Optimization

Chapter 6

Model Classes and Novel Formulations in Finance

In this chapter, we describe existing and develop new optimization problem formulations that are used in financial modeling, in particular in risk management and portfolio selection.

Risk measures, used for quantifying financial risks, and mean-variance portfolio selection are described in Section 6.1. Mean-variance framework assumes normality of distribution of asset returns. Unlike mean-variance optimization, scenario-based approaches, see Sections 6.2 and 6.3, allow for general non-normal, discrete and subjective return/loss distributions. Further, scenario-based approaches allow the modeling of nonlinear financial instruments, such as derivatives and options. Scenario-based approaches are one of the methods in *stochastic optimization.*

Using scenario-based approach, we develop and extend two novel financial optimization frameworks in this chapter. *Portfolio replication* models are investigated in Section 6.2 and a portfolio *credit risk optimization* framework is developed in Section 6.3. The results are mostly based on [77] and [19]. The novelty of our results described in Section 6.2 is in applying regularized optimization and parametric optimization to portfolio replication for insurance liabilities. New optimization formulations for minimizing portfolio credit risk are developed in Section 6.3 and are computationally tested on large data sets in Appendix A.

6.1 Risk Measures and Mean-Variance Portfolio Selection

Financial institutions operating in all markets require competitive risk and portfolio management tools. The goal of the investors is to obtain optimal returns in all market environments when risk is involved in every investment, borrowing, lending and project planning activity. Many different types of risk should be taken into account and there are relationships between them. For instance, credit-risk is connected to other risks faced by financial institutions, including market risk. Moreover, there are many risk measures used for quantitative evaluation of risk including *variance*, *value-at-risk* (VaR) and *expected shortfall* (ES). Expected shortfall is also known as *conditional value-at-risk* (CVaR), and thus we will use those two terms interchangeably. The most famous portfolio management model that involves a risk-return tradeoff is the mean-variance portfolio optimization problem introduced by Markowitz [95]. The conflicting objectives in the Markowitz model are minimizing portfolio variance (risk) and maximizing expected return. Using variance to measure a portfolio's risk, as proposed by Markowitz, is still important for finance both in theory and as a practical benchmark. The Markowitz model and many other risk management models can be formulated and solved as optimization problems. In this chapter, we focus on formulating and solving portfolio management, portfolio replication and portfolio credit risk problems represented as optimization problems. The goal is to develop practical optimization tools for replicating portfolios and managing credit risk, as well as to compare them with Markowitz's mean-variance portfolio optimization model. For instance, the Markowitz framework is poorly suited for credit risk, and practical problems require going beyond it.

6.1.1 Portfolio Selection

A general goal of portfolio optimization and risk management is selecting a portfolio of assets, i.e., stocks, bonds, derivatives, etc., such that a large return (or

some other performance indicator) with a low risk is obtained. Overview of portfolio selection, asset allocation and risk management techniques including quantitative analysis can be found in [128, 119, 30, 117] and many other publications. In this chapter we consider a number of portfolio models for mean-variance optimization, credit risk optimization, and portfolio replication in more detail.

The general framework is the following. Consider assets with random returns that are available for investment. Let a portfolio contain N assets represented by positions u_i, $i = 1, 2, \ldots, N$, where positions are measured in units, or by portfolio weights w_i, $i = 1, 2, \ldots, N$, where weights define a portion of total wealth (or total budget) invested in the corresponding stock. The total portfolio value v_P is computed as

$$v_P = \sum_{i=1}^{N} v_i u_i, \tag{6.1.1}$$

where v_i is the value of asset i. The portfolio weights are defined as the relative asset value

$$w_i = u_i \frac{v_i}{v_P}, \quad i = 1, 2, \ldots, N. \tag{6.1.2}$$

Denoting by r_i the expected market return of a security i, the portfolio return, r_P, satisfies the relation

$$r_P = \sum_{i=1}^{N} w_i r_i = r^T w. \tag{6.1.3}$$

From (6.1.1) and (6.1.2) it follows that

$$\sum_{i=1}^{N} w_i = 1, \tag{6.1.4}$$

which is usually called the portfolio budget constraint.

We call $x = (x_1, x_2, \ldots, x_N)^T$ the decision vector. Note that either portfolio weights w or positions u can be used as the decision vector. Mean-variance portfolio optimization traditionally uses weights w as decision variables, while portfolio replication and credit risk optimization models use positions u. In order to simplify the notation, we denote by x either portfolio weights or positions and specify the units of x for each formulation that we consider later on.

6.1.1.1 The Markowitz Mean-Variance Portfolio Optimization Model

In Section 6.1 the decision variables x_i, $i = 1, \ldots, N$ denote portfolio weights or portion of total funds invested in security i. Vector $r = (r_1, \ldots, r_N)^T$ is the vector of expected returns, while σ_i denotes standard deviation of return of asset i. As the Markowitz model takes correlations between assets into account, we introduce ρ_{ij} as the correlation coefficient of assets i and j returns. Correspondingly, $Q = [\sigma_{ij}]$ is the $N \times N$ variance-covariance (or just covariance) matrix of asset returns with the entries $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$. Matrix Q is symmetric and positive semidefinite, i.e., $x^TQx \geq 0$ for all $x \in \mathbb{R}^N$. In practice r and Q are usually computed from historical data.

Additional assumptions that are commonly made in portfolio selection models include if one has a single time period or multi-periods, the availability of an initial portfolio, and presence of risk-free asset, i.e., cash. All portfolio models that we consider in this thesis are single period models. Portfolio re-balancing refers to the availability of an initial portfolio, while portfolio selection usually implies constructing a portfolio from scratch. The availability of an initial portfolio will be considered in credit risk optimization for comparison purposes. We consider risk-free assets later on when describing a Capital Asset Pricing Model.

Expected return and variance of the portfolio x are computed as:

$$\mathbb{E}[x] = r_P = r_1x_1 + \ldots + r_Nx_N = r^Tx,$$

$$\text{Var}[x] = \sigma_P^2 = \sum_{i,j} \rho_{ij}\sigma_i\sigma_jx_ix_j = x^TQx,$$

with $\rho_{ii} = 1$.

Mean-variance portfolio models, see Markowitz [95, 96], which are based on investor's utility maximization, can be formulated as optimization problems and, more precisely, as parametric QO problems. Let us consider the weighted sum formulation due to Farrar [49] in more detail:

$$\begin{array}{ll} \min\limits_{x} & -\lambda r^Tx + \frac{1}{2}x^TQx \\ \text{s.t.} & Ax = b \\ & x \geq 0, \end{array} \tag{6.1.5}$$

with the constraints denoted by $x \in \mathcal{F}$ for simplicity. The set of admissible or feasible portfolios can be defined as $\mathcal{F} = \{x : Ax = b,\ x \geq 0\}$, where linear constraints on asset holdings of the type $Ax = b$ can include, e.g., a budget constraint $\mathbf{1}^T x = \sum_i x_i = 1$ and bounds on asset holdings; $x \geq 0$ is the no-short-sale constraint.

Markowitz [96] defined a portfolio to be efficient if for some fixed level of expected return no other portfolio gives smaller variance (risk). Equivalently, an efficient portfolio can be defined as the one for which at some fixed level of variance (risk) no other portfolio gives larger expected return. The determination of the efficient portfolio frontier in the Markowitz mean-variance model is equivalent to solving the parametric QO problem (6.1.5).

Parameter $\lambda > 0$ in (6.1.5) is referred to as an investor's risk aversion parameter. Solutions of the optimization problem (6.1.5) for different values of λ trace the so-called efficient frontier in the mean-variance space. In Figure 6.1 we plot the mean-variance efficient frontier for a particular instance of problem (6.1.5) in the expected return – standard deviation coordinates. The mean-variance efficient frontier is known to be the graphical depiction of the Markowitz efficient set of portfolios and represents the boundary of the set of feasible portfolios that have the maximum return for a given level of risk. Portfolios above the frontier cannot be achieved. Figure 6.1 shows the efficient frontier in the mean-standard deviation space in order to be consistent with the existing literature.

The mean-variance optimization problem has a number of alternative formulations. Similarly to tracing Markowitz's efficient frontier with the weighted sum method, we can utilize the ε-constrained formulation for that purpose:

$$\begin{array}{ll} \min\limits_{x} & \frac{1}{2} x^T Q x \\ \text{s.t.} & r^T x \geq \varepsilon \\ & Ax = b \\ & x \geq 0, \end{array} \tag{6.1.6}$$

where the parameter $\varepsilon \in [r_{\min}, r_{\max}]$.

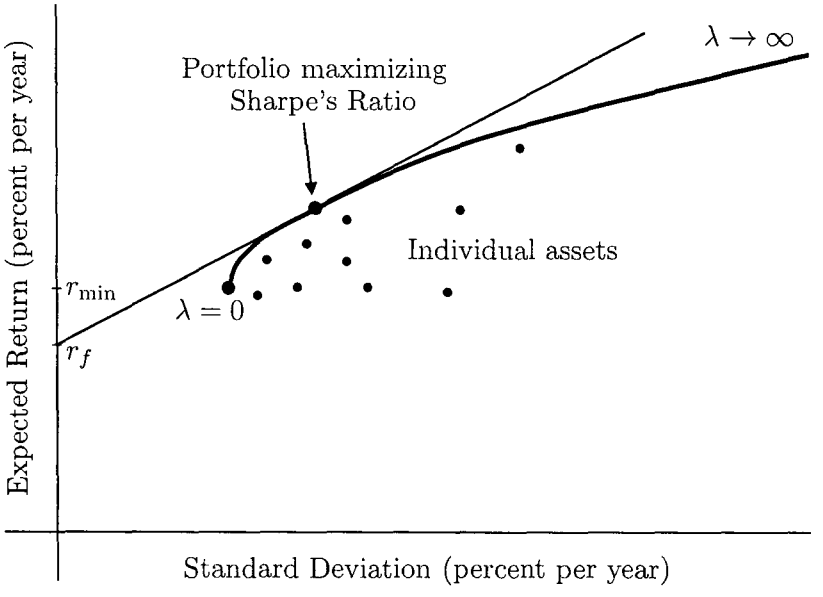


Figure 6.1: Efficient Portfolio Frontier and Capital Market Line.

The third formulation of the mean-variance problem is maximizing the expected return with the ε constraint on the variance. This problem with the variance in the constraint is a parametric QCQO problem. Equivalence between the three mean-variance formulations, namely equations (6.1.5), (6.1.6), and the parametric QOCO problem, is shown in [83]. All the three formulations are multiobjective optimization problems, where the conflicting objectives are maximizing expected return and minimizing variance.

The Markowitz mean-variance model works relatively well in the presence of market risk, when returns are distributed normally, however it is not conforming with reality when other risks, such as credit risk, are present. In Chapter 7 we solve a number of variants of the mean-variance optimization problem and their extensions, including the standard mean-variance optimization problem with market risk in Section 7.2, the mean-variance problem with transaction costs in Section 7.3, and the mean-variance optimization problem with trading penalties in Section 7.5.

6.1.1.2 The Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) [132] states that asset return can be characterized completely by a combination of market return and the asset co-variation with the market [119]. It assumes a linear relationship between the return of an asset and the return of the market, so, it is a single factor model. In addition, CAPM introduces the concepts of specific (or idiosyncratic) risk and systemic (or systematic) risk. Specific risk is unique to an individual asset, while systemic risk is associated with the market. Specific risk can be diversified by holding many different assets in a portfolio.

Suppose that a risk-free asset exists with return r_f. The CAPM relates the random return on the i-th investment, R_i, to the random return on the market R_M by

$$R_i - r_f = \beta_i(R_M - r_f) + \epsilon_i, \tag{6.1.7}$$

where ϵ_i is a random variable with zero mean and σ_{ϵ_i} standard deviation, it is uncorrelated with the market return R_M and $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$; β_i is the *beta* term correlated with the market.

Now, the expected return of asset i is $r_i = \mathbb{E}[R_i]$ and the expected market return is $r_M = \mathbb{E}[R_M]$. Taking the expectation of (6.1.7), we get the CAMP equation:

$$r_i - r_f = \beta_i(r_M - r_f), \tag{6.1.8}$$

with $\mathbb{E}[\epsilon_i] = 0$ and $\mathrm{Cov}(\epsilon_i, R_M) = \mathrm{Cov}(R_i, R_M) - \beta_i \mathrm{Var}(R_M) = \sigma_{iM} - \beta_i \sigma_M^2 = 0$, where σ_M is the standard deviation of market return. Equation (6.1.8) states that the expected excess rate of return on asset i equals to its beta times the expected excess rate of return on the market portfolio.

Now we can compute the standard deviation of the return of asset i as

$$\sigma_i = \mathrm{Var}[R_i] = \beta_i^2 \sigma_M + \sigma_{\epsilon_i}^2 = \beta_i \sigma_{iM} + \sigma_{\epsilon_i}^2. \tag{6.1.9}$$

The first term in equation (6.1.9) reflects systemic risk, while the second term is the idiosyncratic risk. The sum of the two terms is called total risk or variance risk.

It follows that the i-th *asset beta coefficient* can be computed as

$$\beta_i = \frac{\text{Cov}(R_i, R_M)}{\text{Var}(R_M)} = \frac{\sigma_{iM}}{\sigma_M^2}. \tag{6.1.10}$$

Beta is a measure of the degree to which the returns of an asset tend to move with the return of the market, relative to the magnitude of the markets return variation. As a result, beta indicates the risk of an asset relative to the risk of the market, and is referred to as an asset *risk measure*.

For a large portfolio, its expected return is equal to $r_P = r_f + \beta_P(r_M - r_f)$ and the standard deviation of its return is $\sigma_P = |\beta_P|\sigma_M$ with $\beta_P = \sum_i \beta_i x_i$. Coefficient β_P is known as the *portfolio beta*. Variance is the relevant risk measure for a portfolio and assets covariances contribute to portfolio variance. Observe that the contribution from the uncorrelated ϵ's to the portfolio variance vanishes as we increase the number of assets in the portfolio. This is the so-called diversifiable risk. The remaining risk, which is correlated to the market, is the undiversifiable systemic risk.

Within the CAPM framework, minimizing portfolio risk is equivalent to minimizing *portfolio beta* as with beta being small we would expect portfolio performance to be unrelated to the market as a whole. The simplicity of the CAPM makes it attractive, but it may be inadequate as many more factors than just the return on the overall market have an impact on the return of an asset [119].

The Capital Market Line (CML) in the (r_P, σ_P) coordinates is defined as

$$r_P = r_f + \frac{r_M - r_f}{\sigma_M}\sigma_P.$$

In the presence of a risk-free asset, all efficient frontier portfolios lie on the CML and it gives the trade-off between portfolio risk and return.

One can think of the optimal CML as the CML with the largest slope [30]. Mathematically, this can be expressed as the portfolio x that maximizes the quantity

$$\frac{r_P - r_f}{\sigma_P} = \frac{r^T x - r_f}{\sqrt{x^T Q x}}, \tag{6.1.11}$$

subject to relevant constraints $x \in \mathcal{F}$ on asset holdings. Quantity (6.1.11) shows the expected return per unit of risk, and it is known as Sharpe's ratio. Sharpe's ratio is the slope of the CML.

The optimization problem that corresponds to maximizing Sharpe's ratio is:

$$\begin{array}{ll} \max\limits_{x} & \frac{r^T x - r_f}{\sqrt{x^T Q x}} \\ \text{s.t.} & x \in \mathcal{F}. \end{array}$$

In Figure 6.1, CML is a ray starting at $(0, r_f)$. The portfolio that maximizes Sharpe's ratio corresponds to the point where CML is tangential to the Markowitz efficient frontier.

In Section 7.1 we formulate and solve a parametric optimization problem within the CAPM framework.

6.1.1.3 Robust Portfolio Selection

Robust portfolio selection allows reducing sensitivity of the optimal portfolio to data perturbations. First, robust optimization naturally arises in portfolio selection because limited number of scenarios are used to compute cash flows, and those cannot be estimated precisely and thus bare high degree of uncertainty. Robust optimization allows taking that uncertainty into account. Second, parameter estimates and historical data also bare high degree of uncertainty. The following techniques can improve robustness of portfolio selection models [47]:

- Resampling for mean-variance portfolio selection;
- Regularized optimization and trading penalties, see Section 6.2;
- Robust optimization [7].

Regularized optimization allows improving sparsity and stability of optimal portfolios by introducing trading penalty to the objective function. Robust optimization puts uncertain data into a bounded uncertainty set and incorporates that uncertainty set into the optimization formulations.

Let us recall the Markowitz mean-variance model (6.1.5), where the set $\mathcal{F}$ of constraints is known without any uncertainty, but the problem data, namely

expected asset returns and covariances are uncertain. This model faces estimation risk as we use estimates of unknown expected returns and covariances and, consequently, uncertainties in those parameters can produce meaningless portfolios. It creates the need for robust portfolio selection techniques.

For *robust portfolio optimization* we may consider a model where return and covariance matrix information are given in the form of intervals [30]:

$$\mathcal{U} = \{(r, Q) : r^L \leq r \leq r^U,\ Q^L \leq Q \leq Q^U,\ Q \succeq 0\},$$

which, e.g., are generated from historical lows and highs. The robust optimization problem is defined as to minimize the objective function in the worst-case realization of the input parameters r and Q, i.e.,

$$\min_{x \in \mathcal{F}} \left\{ \max_{(r,Q) \in \mathcal{U}} -\lambda r^T x + \frac{1}{2} x^T Q x \right\}.$$

Now, we consider a variant of robust portfolio selection problems proposed by Ceria and Stubbs [22]. In their model, instead of the uncertainty set being given in terms of bounds, they use ellipsoidal uncertainty sets. In [22] the authors assume that only r, the vector of estimated expected returns, is uncertain in the Markowitz model (6.1.5). In order to consider the worst case of problem (6.1.5), it was assumed that the vector of true expected returns r is normally distributed and lies in the ellipsoidal set:

$$(r - \hat{r})^T \Theta^{-1} (r - \hat{r}) \leq \kappa^2,$$

where $\hat{r}$ is an estimate of the expected return, Θ is covariance matrix of the estimates of expected returns with probability η, and $\kappa^2 = \chi_N^2(1 - \eta)$ with χ_N^2 being the inverse cumulative distribution function of the chi-squared distribution with N degrees of freedom.

Let $\hat{x}$ be the optimal portfolio on the estimated frontier for a given target risk level. Then, the worst case (maximal difference between the estimated expected return and the actual expected return) of the estimated expected returns

with the given portfolio $\hat{x}$ can be formulated as:

$$\begin{array}{ll} \max\limits_{\hat{r}-r} & (\hat{r}-r)^T\hat{x} \\ \text{s.t.} & (r-\hat{r})^T\Theta^{-1}(r-\hat{r}) \leq \kappa^2. \end{array} \tag{6.1.12}$$

As derived in [22], by solving problem (6.1.12) we get that the optimal objective value $(\hat{r}-r)^T\hat{x}$ is $\kappa\|\Theta^{1/2}\hat{x}\|_2$. So, the true expected return of the portfolio can be expressed as $r^T\hat{x} = \hat{r}^T\hat{x} - \kappa\|\Theta^{1/2}\hat{x}\|_2$.

Now, problem (6.1.5) becomes a robust portfolio selection problem

$$\begin{array}{ll} \min\limits_{x} & -\lambda\hat{r}^Tx + \frac{1}{2}x^TQx + \kappa\|\Theta^{1/2}x\|_2, \\ \text{s.t.} & x \in \mathcal{F}. \end{array} \tag{6.1.13}$$

Problem (6.1.13) is SOCO problem, moreover, it is a parametric optimization problem with two parameters λ and κ. We will solve an instance of problem (6.1.13) in Section 7.4.

6.1.2 Risk Measures and Quantile-Based Risk Models

The Markowitz model is commonly used in practice in the setting of market risk. From an optimization perspective, minimizing variance requires solving a quadratic optimization problem, which is readily accomplished by commercial or open-source solvers. However, it is well known that variance is a poor measure of risk when the distribution of a portfolio's return is asymmetric. This fact is particularly relevant for hedge funds, for example, which often make extensive use of derivatives. It also applies to portfolios that incur credit risk, namely the potential reduction in the value of a financial instrument (e.g., a bond) due to the default, or the increased likelihood of default of a counterparty to a contract. Instead of measuring risk in terms of variance, managers of credit-risky portfolios typically focus on extreme quantiles of the portfolio's loss distribution. Thus, a credit-risk optimization problem may require minimizing the 99.9 percentile loss, i.e., the loss that is likely to be exceeded with a probability of 0.001, for example. Such problems are typically harder to solve than those that involve variance, as the objective functions (risk measures) are less tractable.

When estimating risks of individual assets and portfolios, risk measures play crucial role. Risk measures are used by practitioners for a number of purposes, including quantifying risk and determining the amount of capital a financial institution needs to hold. A risk measure should be chosen appropriately to serve these purposes. In addition, a risk measure needs to be represented as a tractable objective function if it is meant to be optimized. In this section we discuss a number of risk measures and provide general formulations for the corresponding optimization problems.

The CAPM uses variance (standard deviation) as the risk measure. But as the CAPM does not take correlations between individual assets into account, portfolio variance is just the sum of variances of individual assets. As portfolio variance is proportional to portfolio beta in the CAPM framework, portfolio beta is used as the risk measure. Minimizing risk in CAPM models is equivalent to minimizing portfolio beta $\beta_P(x)$ or just

$$\min_x \sum_i \beta_i x_i.$$

The Markowitz model uses covariance as the risk measure. In many cases the mean-variance risk measure is referred to as portfolio variance. The risk minimization problem in mean-variance framework is minimizing portfolio variance $\mathrm{Var}[x]$:

$$\min_x x^T Q x,$$

where Q is the covariance matrix. If the underlying variables are not normally distributed, variance is not a comprehensive risk measure as it does not capture all relevant risk aspects. Another common criticism of measuring portfolio risk with variance is that the mean-variance model penalizes up-side and down-side risk equally, whereas most investors do not mind taking up-side risk. So, one possible solution is to use quantile-based risk measures.

Our primary interest is in quantile-based risk measures such as value-at-risk (VaR) and expected shortfall (ES or CVaR). VaR is incorporated in the Basel II

Capital Accord, which makes it an integral part of the regulatory requirements. For a detailed review of these risk measures we refer the reader to [138].

Let us consider a return distribution for portfolio x, where the random variable R describes random returns. Then, value-at-risk VaR_α is the largest possible return, at a given confidence level α and over a given period of time, that one will experience [112]:

$$\mathrm{VaR}_\alpha(R) = \max\{\ell : \mathbb{P}(R \leq \ell) \leq 1 - \alpha\} = F_R^{-1}(1 - \alpha).$$

Expected shortfall ES_α is the average return beyond the VaR level:

$$\mathrm{ES}_\alpha(R) = \mathbb{E}[R : R \leq \mathrm{VaR}(R)].$$

In the definitions above, α is the confidence level or quantile, e.g, $\alpha = 0.99$ refers to a 99% confidence level; F_R^{-1} is the inverse of the cumulative return distribution of R. If R describes random payoffs, then VaR is a threshold in terms of monetary units.

Figure 6.2 illustrates risk measures discussed above on the basis of a probability density function of portfolio returns.

If VaR is defined for the distribution of losses L, we get

$$\mathrm{VaR}_\alpha(L) = \ell_\alpha(L) = \min\{\ell : \mathbb{P}(L \leq \ell) \geq \alpha\}.$$

Here, VaR is predicted maximum loss with a specified confidence level over a given period of time.

It is known that VaR has some undesirable properties, namely it does not take into account risks exceeding VaR (for losses), it is not a coherent risk measure [117] as VaR is not sub-additive leading to the fact that portfolio diversification can increase risk, and is non-convex. Expected shortfall is a coherent risk measure and is attractive for optimization due to its convexity.

Let us denote the loss function by f. The loss function f depends on the vector of portfolio weights x that are the control variables, and ξ is the random vector of underlying risk factors with the probability measure $\mathbb{P}$. Risk factors ξ

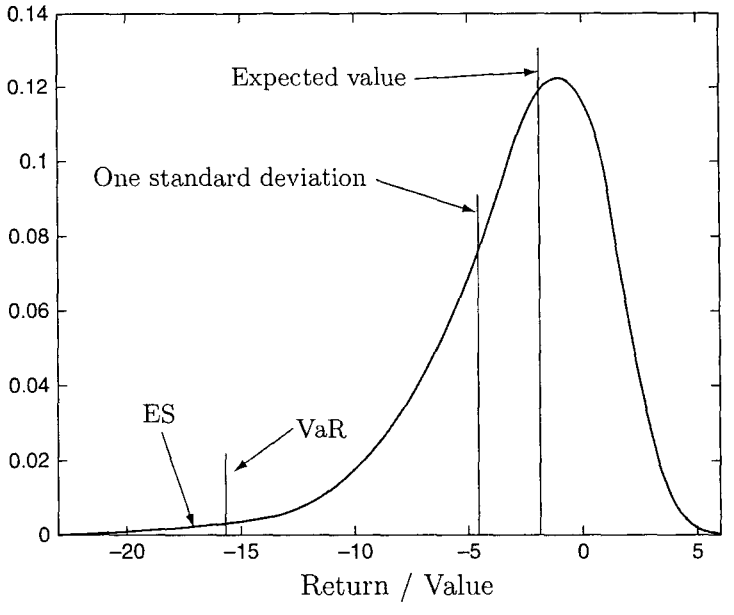


Figure 6.2: Risk Measures (source [112]).

may include random returns R, random losses L, random macroeconomic variables, etc. Now, the loss function is $f(x,\xi)$ and the reward function is $-f(x,\xi)$.

Examples of optimization problems involving expected shortfall risk measure are presented below. They resemble the formulations for mean-variance optimization. Unlike minimizing VaR, expected shortfall minimization is a convex problem if the function of losses $f(x,\xi)$ is convex in x.

$$\begin{array}{ll} \max\limits_{x} & \mathbb{E}[-f(x,\xi)] \\ \text{s.t.} & \mathrm{ES}_\alpha[f(x,\xi)] \le \nu \\ & x \ge 0, \end{array} \qquad \begin{array}{ll} \min\limits_{x} & \mathrm{ES}_\alpha[f(x,\xi)] \\ \text{s.t.} & \mathbb{E}[-f(x,\xi)] \ge \rho \\ & x \ge 0, \end{array} \qquad \begin{array}{ll} \max\limits_{x} & \mathbb{E}[-f(x,\xi)] \\ \text{s.t.} & \mathrm{ES}_{\alpha_1}[f(x,\xi)] \le \nu_1 \\ & \mathrm{ES}_{\alpha_2}[f(x,\xi)] \le \nu_2 \\ & x \ge 0. \end{array}$$

When mean-variance analysis is extended by utilizing quantile-based measures instead of variance, this generalization is called mean-risk analysis [137]. More information about formulating and solving VaR and ES optimization problems in the credit risk context can be found in Section 6.3.

6.2 Portfolio Replication

Financial institutions and insurance companies require competitive risk and portfolio management tools for measuring and managing risks in order to use those results in business decision making. In particular, insurance companies target meeting their liabilities in the presence of different types of risk. In order to quantify risk types across the entire organization, in particular for an insurance company, the replicating portfolios risk management technology can be utilized. The methodology of *portfolio replication* enables insurers to generate a portfolio of standardized assets as a representation for their liabilities in order to calculate market risk measures on a market consistent basis, faster and more transparently than existing methods. This methodology for measuring risk allows calculating risk capital across the entire organization in order to meet regulatory requirements.

A replicating portfolio comprises a set of standard financial assets whose value closely matches that of a liability portfolio under current and future market conditions [17]. If the replication is sufficiently precise and the assets can be priced faster than the liability, which is the case in insurance industry, then the replicating portfolio is a computationally efficient proxy for conducting risk analysis of the liability. Replicating portfolios can be obtained using optimization techniques. The optimization problem behind the portfolio replication is minimizing the mismatches between cash flows representing the liabilities and returns of the replicating portfolio for a number of future time periods in the presence of uncertainty.

Given a set of candidate assets, an optimization problem is formulated to find the asset positions that best match certain characteristics of the liability in some limited set of projected economic scenarios. Intuitively, replicating portfolios rely on the fact that complex liability cash flows can be decomposed into simpler cash flows that correspond to financial instruments. For example, fixed cash flows can be represented as zero coupon bonds, minimum guarantees for variable annuities as puts on market indices, and fixed annuity options as

swaptions with physical settlement [17].

The focus of this section is on formulating and solving portfolio replication models represented as optimization problems, in particular with regularized and parametric optimization techniques. The first key challenge to its practical implementation is defining an appropriate penalty function for measuring mismatches between cash flows to be replicated and a replicating portfolio. The second challenge is the fact that the distribution of cash flows is not known exactly, and its approximation is constructed from a set of scenarios representing possible cash flows of company's liabilities [39, 99]. The last challenge is that in order to avoid overfitting, that deteriorates out-of-sample performance, a constructed replicating portfolio has to be sparse.

A sparse replicating portfolio, that is desirable for practical reasons, need to contain a relatively small number of assets. The smaller the replicating portfolio, the more quickly it can be priced, and the easier it is to interpret in relation to the liability. Moreover, sparse portfolios are often better able to replicate the liability across a broad range of market conditions. The merits of this approach for replication of insurance liabilities are noted in [35] and a genetic algorithm has been proposed for this purpose in [113]. It was shown in [17] that imposing trading restrictions during construction of a replicating portfolio is an effective way to achieve an appropriate level of sparsity. Trading restrictions serve as an approximation for restricting portfolio cardinality. Moreover, imposing trading restrictions is equivalent to imposing penalty on portfolio size, which is know as regularized problem in optimization. *Regularized optimization* problem can be considered as a variant of multiobjective optimization problem, which solution is the trade-off between minimizing the replication mismatch and maximizing sparsity of the replicating portfolio. Similarly to the risk-reward trade-off in portfolio selection, solutions of the portfolio replication problem with trading penalty produces an efficient frontier between the replication error and solution sparsity.

Replication errors between replicated and replicating portfolios can be de-

fined in terms of different norms. If the ℓ_1-norm or the infinity norm is used to model mismatch errors and all the problem constraints are linear, the optimization problem for minimizing the mismatch is a linear optimization problem. The Euclidian norm leads to a quadratic optimization problem. In some cases, only expected underperformance or a weighted combination of the positive and negative performance deviations of a replicating portfolio from the replicated cash flows is penalized leading to a wider class of models [39].

Computational methods to solve portfolio replication problems include continuous linear and quadratic optimization as well as mixed-integer linear and quadratic optimization. Mixed-integer formulation is employed if cardinality constraints are present. Due to large problem sizes, mixed-integer optimization is not always applicable as practical formulations need to be solved in near real-time, but serves as a tool for comparing solution quality of the regularization approach.

In the remainder of this section we first discuss previous studies on portfolio replication in Section 6.2.1. Notation and concepts that are needed for constructing replicating portfolios are introduced in Section 6.2.2. In Section 6.2.3 we describe the use of parametric optimization for solving portfolio replication problems formulated as regularized optimization problems and weighted regularized optimization problems. The novelty of our results is in applying parametric optimization techniques to compute replication portfolios.

6.2.1 Previous Studies

The portfolio replication problem appearing in insurance industry is closely related to *portfolio replication* problem in asset and portfolio management and *benchmark (index) tracking* problem in portfolio selection. In addition, applications of portfolio replication include hedging and pricing in complete and incomplete markets, and portfolio compression [39].

The portfolio replication problem in asset management is known as *dimension-reduction problem* and consists of replicating a large portfolio containing N assets

with a small portfolio containing K assets [101]. The objective in that problem is minimizing the tracking error of the portfolio that is constrained to contain K assets and a market of N assets. It is a combinatorial problem to select K assets from the universe of N. It requires solving the optimization problem, if constraints are present, or at least evaluating the objective function $\binom{N}{K}$ times. One way to handle it, is by solving a mixed-integer optimization problem, however it can be computationally too expensive for large problems. In addition, K is often a decision variable and in many cases the ideal number of securities K in the replicating portfolio is evaluated by considering trade-offs between the quality of replication, e.g., a tracking error, and the transaction costs [101]. Solving the problem for every possible value of K requires $\sum_{K=1}^{N} K\binom{N}{K} = 2^N$ operations, which is prohibitively slow. To reduce computational complexity, in [101] a simple heuristic is described that evaluates the objective function only $\sum_{K=1}^{N} K = \frac{1}{2}N(N+1)$ times. For each value of K starting from $K = N$ the heuristic finds and drops the "worst" security from the portfolio by evaluating the portfolio K times. Dropping each security one by one is performed for each $K = N, N-1, \ldots, 2$. Even though the implementation of the heuristic is quite simple, it may not produce acceptable results and may still be time consuming if the replication problem has constraints and requires solving an optimization problem for each potentially dropped security.

Benchmark (index) tracking problem in portfolio selection consists of tracking a benchmark or index portfolio with another portfolio. There are a number of reasons to create a replicating portfolio that tracks the benchmark. For instance, the benchmark portfolio may not be tradable or may contain a large number of instruments. Tracking error is usually defined as the standard deviation of a replicating portfolio return relative to a benchmark: $\mathrm{TE}(x) = (x - \bar{x})^T Q (x - \bar{x})$, where x_i, $i = 1, \ldots, N$ is the percentage of the portfolio invested in stock i, $\bar{x}$ is the vector of percentage weights of the stocks in the index and Q is a covariance matrix of the stock returns. So, tracking error measures relative volatility of the replicating portfolio in relation to the benchmark. Relations of the tracking error

to other replicating portfolio features can be studied in a number of ways. For instance, in [18] tracking error is expressed as a function of the trading strategy, allowing to construct trade risk profiles, that express the relationships among tracking error, risk contributions, expected returns and trading costs.

The idea that an index tracking portfolio may consist of relatively few assets has the following background – it prevents portfolio from holding very small positions and limits transaction costs and service fees. Consequently an index tracking algorithm that performs tracking error minimization may limit or restrict the total number of assets in the portfolio. An obvious way to introduce the restriction for the total number of assets in a portfolio is a cardinality constraint. *Cardinality* of a vector $x \in \mathbb{R}^N$ is the number of nonzero components in it. If we define $0^0 = 0$, the cardinality of a vector x is the counting function $\text{card}(x) = \sum_{i=1}^{N} x_i^0 = \|x\|_0$, where $\|x\|_0$ is often called *zero norm*[1]. In general, for a real number $p \geq 1$, *p-norm* is defined[2] as $\|x\|_p = \left(\sum_{i=1}^{N} |x_i|^p\right)^{1/p}$. In index tracking problems, the cardinality constraint has the form $\text{card}(x) = \|x\|_0 \leq K$. Instead of inequality constraint, it can be the equality constraint. Those problems are called *cardinality-constrained optimization* problems. The disadvantage of cardinality-constrained optimization problems is that those are no longer continuous and involve integer variables. Exact solution to the cardinality constrained optimization problem can be achieved by using mixed-integer optimization, see, e.g., [24].

In [80] a heuristic technique is described for cardinality-constrained index tracking problems. The investment universe is the S&P 100 index. In the heuristic algorithm from [80], first the tracking error for the portfolio with 100 stocks is minimized and the 80 largest stocks are selected. Next, the portfolio with those 80 stocks is optimized and 60 stocks are selected. The algorithm stops when the portfolio with the required number of stocks is selected, e.g., 20 stocks. The algorithm easily generalizes to removing any number of stocks (even 1) at

[1]Zero norm is not a true norm as it is not positive homogeneous.

[2]While this formula is also valid for $0 < p < 1$, the resulting function does not define a true norm as it violates the triangle inequality.

each stage. When one stock is removed at each stage, the algorithm becomes a simplified version of the heuristic from [101], where each stock is verified one by one before the decision of its removal from the portfolio is made.

In [80] the discontinuous counting function $\sum_{i=1}^{N} x_i^0$ is approximated by continuous but not continuously differentiable function $\sum_{i=1}^{N} |x_i|^p$, where the value of p is selected to be 0.5. Essentially $\sum_{i=1}^{N} |x_i|^p$ taken to the power $1/p$ is the p-norm approximation of the cardinality constraint. Another approach, called graduated non-convexity, is taken in [28]. The authors propose approximating the discontinuous function x_i^0 by a sequence of continuously differentiable non-convex piecewise quadratic functions that approach x_i^0 in the limit.

Approximating the zero norm cardinality constraint $\sum_{i=1}^{N} x_i^0 \leq K$ with a p-norm constraint $\sum_{i=1}^{N} |x_i|^p \leq \varepsilon$ seems very attractive. Parameter ε is varied for tuning the cardinality approximation to match K non-zeros in the solution. In addition, the constraints obtained by the ℓ_1-norm for $p = 1$, and the Euclidian norm for $p = 2$ in the form $\|x\|_1 \leq \varepsilon$ and $\|x\|_2 \leq \varepsilon$ are convex constraints. The convex optimization problems that approximate the cardinality constraint can be solved efficiently both in theory and practice. These techniques are known as *regularized optimization*, in particular ℓ_1-regularization refers to the ℓ_1-norm approximation, and Tikhonov regularization refers to the Euclidian norm approximation of the cardinality constraint.

It is important to recognize that the problem of asset selection in portfolio replication is identical to the subset selection problem in linear regression, see Section 5.2, except for the absence of the intercept term. The subset selection problem deals with selecting an appropriate set of predictor variables from a large number of candidates. Like replicating portfolios, statistical models benefit from sparsity. When a model includes only the most important predictors interpretation is easier and overfitting is less likely to occur, and it produces a more stable model with improved out-of-sample prediction accuracy [17].

The methods that have been proposed for variable selection in statistics include those that also employ regularization, see the survey [68]. In this case,

the usual objective of minimizing the regression residuals is augmented with a constraint or a penalty term that discourages the inclusion of unnecessary predictors. Examples of methods that use this technique are the LASSO [140] for ordinary least squares regression and the LAD-LASSO [145] for least absolute deviation regression.

Regularization techniques have been also explored in the context of financial optimization. It is well known that in mean-variance optimization imprecise estimation of the sample's means and covariances can produce unstable portfolios that perform poorly out-of-sample. In order to mitigate these effects, in [16, 103] regularization is incorporated in the form of penalties and constraints on the portfolio weights. A similar approach is used in [62] when minimizing the value-at-risk and the conditional value-at-risk of a portfolio.

In most of the previous studies replication error of the replicating portfolio is considered in relation with other portfolio performance indicators, e.g., portfolio return or transaction cost of constructing that portfolio. Presence of those performance indicators allow selecting the size (sparsity) of the replicating portfolio. In insurance portfolio replication, the difficulty is that there are no other obvious performance indicators except for the mismatch between the liabilities cash flows and cash flows generated by the replicating portfolio. In addition, there are no transaction costs associated with constructing replicating portfolio as it is essentially a virtual portfolio. One way is to test out-of-sample performance of the replicating portfolio, but this tactic requires a systematic way to obtain out-of-sample scenarios for cash flows of liabilities, that may be difficult to generate due to actuarial systems used by insurance companies. Without out-of-sample testing, cross validation can be used on the set of existing scenarios.

6.2.2 Construction of Replicating Portfolios

We target replicating the cash flows of liabilities with a set of N candidate replicating instruments. According to actuarial practice, we assume that cash flows occur at times $t = 1, 2, \ldots, T$ with the convention that $t = 0$ is the present

time and $t = T$ is the time horizon. Uncertainty about future cash flows is described by a set of S scenarios. Each scenario describes the liability and instrument cash flows under different economic conditions and usually includes risk-neutral and real-world scenarios. The per-unit cash flow of instrument i ($i = 0$ for the liability) in scenario l at time t is denoted by c_{li}^t.

Variables x_i, $i = 1, \ldots, N$ denote the position size in units of instrument i in the replicating portfolio. The goal of the portfolio replication optimization problem is to determine the decision vector $x = (x_1, \ldots, x_N)^T$ such that the replicating portfolio matches, as closely as possible, some characteristics of the liability cash flows [17]. Examples of objectives that should be matched in each scenario, among others, include:

- The present values of the cash flows;
- The terminal (accrued) values of the cash flows;
- The cash flows at every time step;
- A set of time-bucketed cash flows.

Typically, a replicating portfolio is obtained by minimizing some measure of the discrepancy in the chosen characteristics. Let $f(x)$ denote such a measure.

A popular measure, motivated by least squares regression, is the weighted sum of squared differences. For example, matching the values of cash flows with the *average square error* measure implies

$$f(x) = \sum_{l=1}^{S} \sum_{t=1}^{T} \omega_l^t \varrho_l^t \left(\sum_{i=1}^{N} c_{li}^t x_i - c_{l0}^t \right)^2, \tag{6.2.1}$$

where ϱ_l^t is the time t discount factor in scenario l, and ω is a set of non-negative weights that prioritizes the liability cash flows, i.e., the larger ω_l^t, the more important is to match c_{l0}^t. Usually, weight $\omega_l^t = p_l$, where p_l is the probability of scenario l.

When matching the cash flows at *every time step*, we usually choose $\varrho_l^t = 1$

and then formula (6.2.1) simplifies to:

$$f(x) = \sum_{l,t}^{S \cdot T} p_l \left(\sum_{i=1}^{N} c_{li}^t x_i - c_{l0}^t \right)^2 . \tag{6.2.2}$$

When matching the *present values* of cash flows, formulation (6.2.1) can be rewritten as

$$f(x) = \sum_{l=1}^{S} p_l \left(\sum_{i=1}^{N} \left(\sum_{t=1}^{T} \varrho_l^t c_{li}^t \right) x_i - \left(\sum_{t=1}^{T} \varrho_l^t c_{l0}^t \right) \right)^2 , \tag{6.2.3}$$

where ϱ_l^t is computed according to some predefined discounting schema.

We can write formulations (6.2.1)-(6.2.3) in a general form:

$$f(x) = \|W(Cx - c_0)\|_2^2 = \|Ax - b\|_2^2, \tag{6.2.4}$$

where W is the diagonal matrix with its diagonal elements being equal to $\sqrt{p_l}$; the elements of matrix C are equal to $C_{ji} = c_{li}^t$, $j = l \times t = 1, \ldots, S \cdot T$ for matching cash flows at every time step or $C_{li} = \sum_{t=1}^{T} \varrho_l^t c_{li}^t$, $l = 1, \ldots, S$ for present value matching; vector c_0 consists of the liability cash flows; matrix $A = W \cdot C$ contains weighted cash flows of the instruments; and $b = W \cdot c_0$ is the vector of weighted liability cash flows.

Another measure, consistent with least absolute deviation regression, is the weighted sum of the absolute differences. In this case, matching the values of cash flows with the *average absolute error* measure results in

$$f(x) = \sum_{l=1}^{S} \sum_{t=1}^{T} p_l \varrho_l^t \left| \sum_{i=1}^{N} c_{li}^t x_i - c_{l0}^t \right| , \tag{6.2.5}$$

that can be rewritten similarly to (6.2.4) with the norm notation as:

$$f(x) = \|W(Cx - c_0)\|_1 = \|Ax - b\|_1. \tag{6.2.6}$$

The dimension of matrix A and vector b in formulations (6.2.4) and (6.2.6) depends on the objectives that should be matched in each scenario. For matching

present values or future (terminal) values of cash flows the dimension of matrix A is $S \times N$, which is essentially equivalent to one time step matching. Using time-bucketed cash flows also results in reducing problem dimensionality by reducing the number of time periods from T to T_b, where $T_b < T$. The most general setting is matching the cash flows at every time step. In that case, matrix A in optimization problems (6.2.4) and (6.2.6) has the largest dimensionality of $T \cdot S \times N$. Choice between those matching strategies depends on the out-of-sample performance of the replicating portfolio and is out of the scope of this thesis.

Now, the general formulation of portfolio replication problem $\min_x f(x)$ without trading penalty is

$$\begin{array}{ll} \min\limits_{x} & \|Ax - b\| \\ \text{s.t.} & x \in \Omega, \end{array} \tag{6.2.7}$$

where the norm $\|\cdot\|$ is either ℓ_1-norm or ℓ_2-norm and we allow for linear equality and inequality constrains $x \in \Omega$ in the formulation.

Note that formulation (6.2.7) is known in multicriteria optimization literature as goal programming. Let us write ℓ_1-norm minimization problem (6.2.7) as:

$$\begin{array}{ll} \min\limits_{x,y} & \sum\limits_{l=1}^{S}(w_l^+ y_l^+ + w_l^- y_l^-) \\ \text{s.t.} & \sum\limits_{i=1}^{N} A_{li}x_i - y_l^+ + y_l^- = b_l,\ l = 1, \ldots, S \\ & y_l^- \geq 0,\ y_l^+ \geq 0,\ l = 1, \ldots, S \\ & x \in \Omega, \end{array} \tag{6.2.8}$$

where variables y_l^+ and y_l^- represent positive and negative mismatches in scenario l and weights w_l^+, w_l^- penalize positive and negative deviations differently. For portfolio replication problems, usually $w_l = w_l^+ = w_l^- = p_l$.

Formulation (6.2.8) is a *goal programming* problem [102, 136]. Goal programming is a technique for solving multicriteria optimization problems, where ideal values of the objectives are known. The goal programming formulation (6.2.8) is a LO problem. In the portfolio replication problem the goal is to match

cash flows in each scenario and this represents multiple conflicting criteria.

If the goal is to penalize squared mismatches $\|Ax-b\|_2^2$, we get the following formulation:

$$\begin{aligned} \min_{x,y} \quad & \frac{1}{2}\sum_{l=1}^{S}[w_l^+(y_l^+)^2 + w_l^-(y_l^-)^2] \\ \text{s.t.} \quad & \sum_{i=1}^{N} A_{li}x_i - y_l^+ + y_l^- = b_l,\ l = 1,\ldots,S \\ & y_l^- \geq 0,\ y_l^+ \geq 0,\ l = 1,\ldots,S \\ & x \in \Omega, \end{aligned} \tag{6.2.9}$$

which is a QO problem. Problem (6.2.9) is the least squares minimization problem (with additional constraints $x \in \Omega$) where we minimize the squared Euclidian norm of mismatches.

6.2.3 Portfolio Replication via Parametric Optimization

As we already know, solving the unconstrained problem $\min_x f(x)$, where $f(x)$ is of the type (6.2.4) or (6.2.6), often produces a replicating portfolio with non-zero x_i for most i, essentially overfitting the liability cash flows in the set of S scenarios. However, we can regularize problem (6.2.7) as described in Section 6.2.1 to reduce the number of non-zero x_i components. Replicating portfolio with "small" x is known to be less sensitive to errors in matrix A containing the scenario cash flows, than a portfolio with large x [15]. So, it is expected that replicating portfolio with relatively small number of instruments may perform well out-of-sample. As a result, limiting the number of instruments in the portfolio allows avoiding overfitting.

In order to minimize the mismatches and limit the number of non-zeros in x, we may introduce the cardinality constraint in the form $\text{card}(x) \leq K$, where K is the desired number of instruments. As we know from Section 6.2.1, the cardinality constraint can be approximated via regularization techniques, where the regularization term that allows limiting the size of x, and, thus, limiting its cardinality, is introduced into the objective. Regularization terms that are

commonly used in practice are $\|x\|_1$ and less frequently $\|x\|_2$.

If ℓ_1-norm regularization is used, then the optimization problem is

$$\min_x \|Ax - b\|_2^2 + \lambda\|x\|_1, \tag{6.2.10}$$

or

$$\min_x \|Ax - b\|_1 + \lambda\|x\|_1, \tag{6.2.11}$$

where $\lambda > 0$ and the resulting problem is called ℓ_1-norm *regularized optimization*.

If the number of scenarios S is smaller then the number of instruments N in the portfolio, we can solve the problem:

$$\begin{array}{ll} \min\limits_x & \|x\|_1 \\ \text{s.t.} & Ax = b, \end{array}$$

which is known as *basis pursuit* criterion [15] and is the special case of the formulation with the constraint $\|Ax - b\| \leq \varepsilon$, $\varepsilon = 0$.

Least squares problem with ℓ_1-norm regularization (6.2.10) can be transformed to:

$$\begin{array}{ll} \min\limits_x & \|Ax - b\|_2^2 \\ \text{s.t.} & \|x\|_1 \leq \varepsilon, \end{array} \tag{6.2.12}$$

which is known as the least absolute shrinkage and selection operator (LASSO), see, e.g., [140, 9]. In (6.2.12) the parameter $\varepsilon \geq 0$ is used to parameterize the importance of sparse solutions. For $\varepsilon = 0$, the optimal solution of (6.2.12) is $x^* = 0$. When $\varepsilon \geq \|x_{LS}^*\|_1$, where x_{LS}^* is the (unconstrained) least-squares solution, we get $x^* = x_{LS}^*$. More detailed overview of ℓ_1 penalized regressions (6.2.10) and (6.2.12) can be found in [68] and [145].

Let us define a generic "cost" of trading as in [17] by

$$h(d, x) = \sum_{i=1}^{N} d_i|x_i| = \|Dx\|_1, \tag{6.2.13}$$

where $d_i > 0$, $d = (d_1, \ldots, d_N)^T$, and D is the diagonal matrix containing weights $d_1, \ldots, d_N$ on its diagonal, i.e., $D = \text{diag}(d)$.

Problems (6.2.10)-(6.2.12) can be generalized further by replacing the regularization term $\|x\|_1$ with the generic cost of trading term $\|Dx\|_1$. In that case, generalized problems (6.2.10) and (6.2.12) are said to include a trading budget and a trading penalty, respectively. This generalization is known as *weighted ℓ_1 regularized* optimization. There are a number of techniques for computing matrix D. The one described in [20] is based on first solving the problem $\min_x \|Ax - b\|_2^2$ without the regularization term $\|Dx\|_1$ and obtaining the optimal solution x^*. After that, the weights are computed as $d_i = \frac{1}{|x_i^*|}$, $i = 1, \ldots, N$ and the weighted ℓ_1 regularization problem is solved.

As we allow for linear constraints on the variables x, we are interested in solving problems (6.2.10) and (6.2.12) with additional constraints $x \in \Omega$. Allowing for linear constraints $x \in \Omega$ in weighted regularized optimization, we can write our formulation of interest based on (6.2.12) as:

$$\begin{array}{ll} \min\limits_{x} & \|Ax - b\|_2^2 \\ \text{s.t.} & \|Dx\|_1 \leq \varepsilon \\ & x \in \Omega. \end{array} \tag{6.2.14}$$

Optimization problem (6.2.14), that we are targeting to solve, is a multiobjective optimization problem, where the conflicting objectives are minimizing replication error and maximizing solution sparsity. As Chapters 2 and 3 explain, we can solve multiobjective optimization problems of the form (6.2.14), including problems with $\|Ax - b\|_1$ objective, as well as (6.2.10)-(6.2.11) using parametric QO techniques.

6.3 Portfolio Credit Risk Optimization

For financial institutions, the benefits of managing (portfolio) credit risk include not only reduced monetary losses due to defaulted or downgraded obligations but also lower capital charges. While individual credit-risky positions can be hedged with credit derivatives such as credit default swaps, imperfectly correlated credit movements among counterparties also provide opportunities for mitigating credit

risk at the portfolio level through diversification. In particular, using optimization techniques to restructure portfolios of credit-risky positions is an attractive possibility. However, such procedures face numerous challenges, foremost being the difficulty of representing the portfolio credit loss distribution with sufficient accuracy. We formulate several alternative optimization problems that are derived from a structural Merton model of portfolio credit risk, and evaluate their effectiveness from the perspectives of risk mitigation and computational practicality.

Credit risk refers to the potential monetary loss arising from the default, or a change in the perceived likelihood of default, of a counterparty to a financial contract. Note that a reduction in the default probability, i.e., a transition to a more favourable credit state, results in a monetary gain. However, such gains are generally small relative to the losses that occur due to severe credit downgrades or default. Thus, the credit loss distribution (F) for a typical investment-grade portfolio is positively skewed, the long right tail being consistent with a small likelihood of substantial losses.

The complex relationships among asset prices, exposures and credit transitions preclude obtaining a closed-form representation of the actual credit loss distribution. Thus, for risk management purposes, it is necessary to replace F by some approximating distribution $\hat{F}$. The form of $\hat{F}$ varies depending on the underlying credit loss model. For example, reduced-form models, e.g., CreditRisk+ [33], provide $\hat{F}$ in closed form. However, their underlying assumptions may be viewed as overly simplistic in that they fail to capture the effects of credit state migrations and correlated movements of risk factors [37]. In contrast, structural models [67, 78] can provide a more realistic representation but typically require $\hat{F}$ to be an empirical distribution derived from Monte Carlo (MC) simulation. For example, the empirical distribution $\hat{F}$ of portfolio losses due to credit events, which is obtained by simulating the portfolio under a set of possible future outcomes (scenarios), is shown in Figure 6.3.

Computing $\hat{F}$ from Monte Carlo simulation presents challenges for assess-

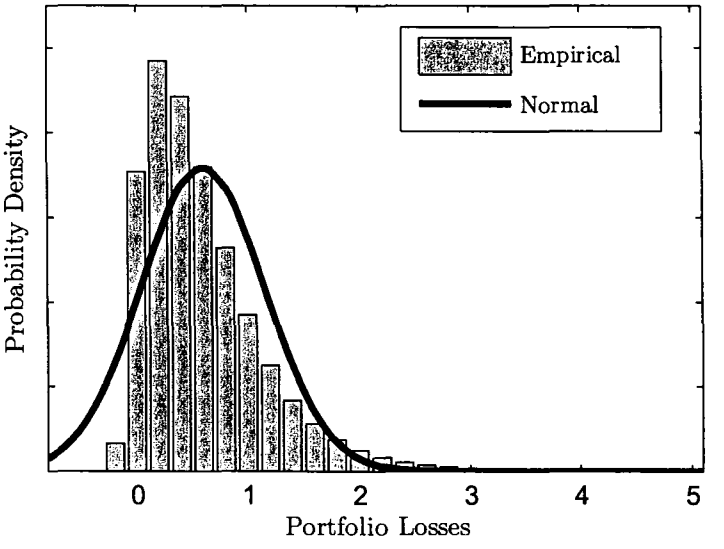


Figure 6.3: Credit Portfolio Loss Distributions.

ing credit risk because common risk measures, such as value-at-risk (VaR) and expected shortfall (ES), involve extreme quantiles in the right tail. Thus, obtaining accurate risk estimates requires a huge number of samples, or scenarios. Initial attempts at minimizing credit risk relied exclusively on MC simulation and included the full set of loss scenarios in the formulation [97, 98, 5, 158, 156]. Clearly, a limitation of this approach is that the large size of the resulting optimization problem adversely affects computational performance. Subsequently, in [130] a large-portfolio approximation was used to obtain a more compact formulation.

More recently, variance-reduction techniques such as importance sampling have been shown [141] to provide stable optimal solutions, with a relatively small number of scenarios. However, a potential problem with importance sampling, is that the required shift in distribution depends on the portfolio's risk, which of course changes with the portfolio's composition during the course of the optimization. Thus, it is not clear that the shift induced by the initial portfolio is also effective for the optimal portfolio.

Structural models infer a counterparty's credit state from its associated creditworthiness index, which depends on systemic risk factors in the form of credit drivers as well as a specific risk factor unique to each counterparty [78, 75]. Given a set of values for the credit drivers, credit transitions for all counterparties become independent. This conditional independence property can be exploited to obtain $\hat{F}$ in semi-analytical form, specifically, as a mixture of closed-form conditional loss distributions. Such representations are far more data-efficient than pure Monte Carlo sampling and the associated optimization problems are smaller as a result. We evaluate the practicality of optimizing credit risk for three different representations of $\hat{F}$:

- Monte Carlo sampling;
- A mixture of normal (Gaussian) conditional loss distributions;
- A mixture of conditional mean (expected) losses.

For comparison purposes, we also consider the performance of variance-based formulations as a way of reducing a portfolio's VaR and ES. Variance minimization, which dates back to the seminal work of Markowitz [95], remains in widespread use for risk management purposes. It is well known that minimizing variance has the effect of also minimizing VaR and ES only for normal distributions. Thus, in our context, minimizing variance effectively assumes that $\hat{F}$ is normal. This is likely to be a poor approximation to the actual portfolio credit loss distribution as Figure 6.3 shows. Nevertheless, its popularity makes variance minimization a useful benchmark when evaluating the performance of the structurally based formulations. Since variance only measures dispersion around the mean, as a second, related benchmark we also minimize the second moment of the credit loss distribution which takes the mean into account.

Our formulations and computational experiments are intended to be consistent with managing the risk of a banking book. Since a typical banking book may contain thousands of counterparties, we allow for optimizing over groups of counterparties. Thus, a portfolio manager might elect to assign all counterparties from a given industry to the same group, for example, and then use the results

of the optimization to restructure the portfolio at the industry level. Such an approach is much more practical than enacting changes to a large number of individual contracts, as might be suggested by optimizing at the counterparty level. We also limit the amount of trading to what can be implemented reasonably when rebalancing the banking book; namely short positions are not permitted and new groups may not be added to the existing portfolio. These limitations are enforced only to provide a realistic assessment of the optimization results; nothing precludes relaxing or eliminating such restrictions from a formulational standpoint. Finally, although we account for credit migration, it is assumed that exposures are deterministic, i.e., positions are not marked to market.

The rest of this chapter is organized as follows. In Section 6.3.1 we introduce some notation and basic concepts, and describe relevant input data for modelling credit-risky instruments. Section 6.3.2 introduces the structural model for portfolio credit risk, for which future credit events may be simulated. In Section 6.3.3, several approximations are described for the loss distribution, F. Section 6.3.4 reviews the risk measures that will be optimized. The formulations of our credit-risk optimization problems follow in Section 6.3.5. We evaluate and analyze our computational results in Appendix A.

6.3.1 Portfolio Credit Losses

We are concerned with credit-risk modelling and optimization at the portfolio level only. Counterparty-level data is used as input to the portfolio-level models. Such data may be estimated from an internal model or provided by an external agency. We start by analyzing the input data for our portfolio level credit-risk optimization problems. Table 6.1 summarizes available information about counterparty-level credit risk.

We consider a single time period. At the end of the time period, each counterparty can migrate to a different credit state resulting in our losses ℓ_c^j, where $j = 1, \ldots, N_{\text{CP}}$ indexes counterparties and c indexes credit states. There are C credit states available for counterparties, enumerated from $c = 0$ (de-

Table 6.1: Credit Instruments Data.

Counterparty		Credit Driver			CS	Recovery after Default		Exposure	Market Return
Group #	CP ID	Credit Driver	CDIDX	Sensitivity Vector	Credit State	Recovery Mean	Recovery Standard Dev	Value	E[Return]
1	USBUSIN1234	DJUSBM	1	0.5568	A	0.62	0.3567	10,255,741	7.0315%
1	USPAPER1234	DJUSBM	1	0.5568	A	0.6	0.3600	9,997,200	7.0435%
129	FRANSTEEL1234	E2BSC	21	0.5399	BB	0.7	0.3368	49,911,974	10.4248%
203	JAPAHOTEL1234	P1CYC	32	0.5154	CCC	0.62	0.3567	5,466,820	36.1575%
271	AUSTREAL1234	P2FIN	44	0.5380	BBB	0.66	0.3481	4,993,662	8.2542%

Counterparty		Credit State Migration Probabilities							
Group #	CP ID	Pr(Default)	Pr(CCC)	Pr(B)	Pr(BB)	Pr(BBB)	Pr(A)	Pr(AA)	Pr(AAA)
1	USBUSIN1234	0.0002	0.0001	0.0013	0.005	0.0507	0.9183	0.0239	0.0005
1	USPAPER1234	0.0002	0.0001	0.0013	0.005	0.0507	0.9183	0.0239	0.0005
129	FRANSTEEL1234	0.0127	0.0055	0.0705	0.8512	0.0545	0.005	0.0005	0.0001
203	JAPAHOTEL1234	0.255	0.68	0.0418	0.0174	0.0058	0	0	0
271	AUSTREAL1234	0.0018	0.0016	0.008	0.0488	0.8849	0.052	0.0024	0.0005

Counterparty		Losses in Credit States							
Group #	CP ID	L(Default)	L(CCC)	L(B)	L(BB)	L(BBB)	L(A)	L(AA)	L(AAA)
1	USBUSIN1234	10,255,741	503,640	201,280	59,625	11,484	0	-500	-1,500
1	USPAPER1234	9,997,200	939,798	398,956	124,925	23,465	0	-3,250	-5,000
129	FRANSTEEL1234	49,911,974	3,571,706	1,283,666	0	-533,837	-677,873	-698,786	-710,178
203	JAPAHOTEL1234	5,466,820	0	-230,494	-375,416	-440,986	-458,134	-461,911	-463,096
271	AUSTREAL1234	4,993,662	328,776	169,824	54,867	0	-15,392	-18,599	-19,623

fault) through increasing credit ratings, to the highest credit rating $c = C - 1$. Note that negative losses (gains) are incurred if a counterparty migrates to a more favourable state. The probability of being in the state c at the end of the time period is p_c^j with p_0^j being the probability of default. Unadjusted exposure of counterparties at default corresponds to their values v_j^{CP} ($v_j^{\mathrm{CP}} > 0$). In general, counterparty exposure is the economic loss that will be incurred on all outstanding transactions if a counterparty defaults, unadjusted by possible future recoveries. Exposures are computed subject to netting, mitigation and collateral. The recovery at default is assumed to be deterministic and recovery-adjusted exposures are equal to $v_j^{\mathrm{CP}}(1 - \gamma_j)$, where γ_j is the recovery rate.

A portfolio consists of N_{CP} counterparties grouped into N_{G} groups, see portfolio snapshot in Table 6.1. For instance, counterparties from the same country, the same industry and having the same credit rating can be grouped together. Changing positions in groups is more practical as it only allows altering

the between-groups positions, leaving the problem of within-group rebalancing as a lower level problem. From the point of view of modelling and optimization, grouping decreases the number of decision variables. Our modelling assumptions do not place any restrictions on the number of groups, group sizes or group compositions.

The value of the i-th group G_i, is $v_i = \sum_{j \in G_i} v_j^{\mathrm{CP}}$, and its loss is

$$\mathcal{L}_i = \sum_{j \in G_i} \sum_{c=0}^{C-1} \ell_c^j \cdot \mathbf{1}\{\text{CP } j \text{ is in credit state } c\}, \tag{6.3.1}$$

where $\mathbf{1}\{\}$ is the indicator function of the event in braces.

Let the decision variable, $x_i \geq 0$, $i = 1, 2, \ldots, N_{\mathrm{G}}$, denote the position in the i-th group, and $x = (x_1, x_2, \ldots, x_{N_{\mathrm{G}}})^T$. Let x^0 be the vector of positions in the initial portfolio. We set $x_i^0 = 1$ for all i, so that the positions are expressed as multiples of the initial holdings. The portfolio value is

$$v(x) = \sum_{i=1}^{N_{\mathrm{G}}} v_i x_i. \tag{6.3.2}$$

The initial portfolio value is then $v_P = \sum_{i=1}^{N_{\mathrm{G}}} v_i$. The portfolio loss $\mathcal{L} \equiv \mathcal{L}(x)$ is defined as

$$\mathcal{L}(x) = \sum_{i=1}^{N_{\mathrm{G}}} \mathcal{L}_i x_i. \tag{6.3.3}$$

6.3.2 Structural Model for Portfolio Credit Risk

The empirical distribution $\hat{F}$ of portfolio losses due to credit events, is obtained by simulation of an underlying structural model described below. Our simulation approach is based on the CreditMetrics framework [67, 129] and the credit-risk portfolio framework from [78, 75]. The degree to which $\hat{F}$ approximates the true distribution F, and thus the quality of the associated risk estimates, depends on the number of samples. The effect of sample size is especially pronounced when estimating quantiles close to 1, as those lie in the extreme right tail of the

distribution. It is well known that the variability of a (risk) estimate decreases as the number of samples increases.

A particular value of portfolio loss ℓ, is computed as a function of the sampled values of a set of risk factors, see (6.3.6), that can be separated into two groups:

- Y denotes a set of systemic risk factors: credit drivers which are macroeconomic factors and sector indices;
- Z denotes a random vector of counterparty-specific, or idiosyncratic, credit-risk factors.

The joint distribution of default and migration events is described through the counterparties' *creditworthiness indices*. The creditworthiness index W_j determines the financial health of counterparty j, and is defined as

$$W_j = \beta^j Y_{n(j)} + \sqrt{1-(\beta^j)^2} Z_j, \tag{6.3.4}$$

where Z_j is the idiosyncratic risk which is independent across counterparties and is normally distributed, $\mathcal{N}(0,1)$; $Y_{n(j)}$ is the counterparty's credit-driver, a standard normal random variable; credit drivers are correlated and normally distributed, $\mathcal{N}(0,\mathcal{C})$ with a given correlation matrix, $\mathcal{C}$; β^j is the factor loading parameter or the sensitivity of the counterparty j to its credit driver $Y_{n(j)}$. As a result, the creditworthiness index W_j is normally distributed, $\mathcal{N}(0,1)$. Thus we are assuming that the creditworthiness index for each counterparty depends on one credit driver; i.e., the counterparty participates in only one sector (e.g., country-industry pair).

$P_c^j = \sum_{\varsigma \le c} p_\varsigma^j$ is the cumulative probability of counterparty j being in credit state c or lower, so that $p_c^j = P_c^j - P_{c-1}^j$, with the convention, $P_{-1}^j \equiv 0$. The credit-state boundaries $\{B_c^j\}_{c=0}^{C-2}$ are defined as $B_c^j = \Phi^{-1}(P_c^j)$, $0 \le c \le C-2$ ($B_{-1}^j \equiv -\infty$, $B_{C-1}^j \equiv \infty$) due to $\mathbb{P}(W_j < B_c^j) = P_c^j$. A counterparty j is in credit state c at the time horizon if $B_{c-1}^j \le W_j < B_c^j$, or equivalently:

$$\mathbb{P}(\text{CP } j \text{ is in credit state } c) = \mathbb{P}(B_{c-1}^j \le W_j < B_c^j). \tag{6.3.5}$$

For this model, the group loss in (6.3.1) takes the specific form,

$$\mathcal{L}_i = \sum_{j \in G_i} \sum_{c=0}^{C-1} \ell_c^j \cdot \mathbf{1}\{B_{c-1}^j \leq \beta^j Y_{n(j)} + \sqrt{1-(\beta^j)^2} Z_j < B_c^j\}. \tag{6.3.6}$$

The key property of this model is *conditional independence*: given a value y of the credit drivers, Y, the creditworthiness indices are independent. Conditional independence allows us, in principle, to obtain the conditional loss distribution by convolution using Fast Fourier Transform. The downside of the convolution technique is that it is difficult to use it for optimization because of the large number of possible losses at the portfolio level. In the next section, we look at some practical alternatives.

6.3.3 Loss Distribution Approximations

Conditional independence gives rise to several variants of the credit-loss distribution model. Scenarios on Y are generated and then the *conditional* loss distribution is approximated by one of the methods. The *unconditional* loss distribution $\hat{F}$ is obtained as the mixture of the conditional loss distributions. Methods for approximating conditional loss distributions [76] include MC-sampling approximation, see Section 6.3.3.1; Cental Limit Theorem (CLT) approximation, see Section 6.3.3.2; and Law of Large Numbers (LLN) approximation, see Section 6.3.3.3.

The number of scenarios can be greatly reduced if one makes simplifying assumptions about the loss distribution and/or the portfolio's composition. For example, in [130] an LLN approximation is also used in which the number of issuers is so large that their individual risks effectively "cancel out" statistically, i.e., the idiosyncratic risk is eliminated.

6.3.3.1 MC-Sampling Approximation

If we generate a sample y from the distribution of Y, the creditworthiness indices are conditionally independent given $Y = y$. Since the idiosyncratic credit-risk

factors are independent of the credit drivers, any number of samples z can be combined with the sample y while still preserving the required codependence structure.

Under the MC-sampling approximation, MK scenarios are generated as follows:

1. Generate a random sample y_l, $l = 1, \ldots, M$ of systemic factors from the distribution of Y.

2. For each $l \in \{1, 2, \ldots, M\}$, generate a random sample z_{lk}, $k = 1, \ldots, K$ of idiosyncratic factors from the distribution of Z (independently, across l).

We denote the j-th counterparty's loss in the (l, k)-th scenario, (y_l, z_{lk}), by

$$\ell_{lk}^j := \sum_{c=0}^{C-1} \ell_c^j \cdot \mathbf{1}\{B_{c-1}^j \leq \beta^j (y_l)_{n(j)} + \sqrt{1 - (\beta^j)^2}\, (z_{lk})_j < B_c^j\}, \tag{6.3.7}$$

so that the sampled value of the i-th group's loss is (cf. (6.3.6))

$$\ell_{i,lk} = \sum_{j \in G_i} \ell_{lk}^j.$$

The sampled portfolio loss is

$$L_{lk}(x) = \sum_{i=1}^{N_G} \ell_{i,lk} x_i.$$

The MC-sampling approximation to F, the cumulative distribution function (cdf), of the portfolio losses, is computed as

$$\hat{F}^{\mathrm{MC}}(\ell; x) = \frac{1}{MK} \sum_{l,k} \mathbf{1}\{L_{lk}(x) \leq \ell\}. \tag{6.3.8}$$

6.3.3.2 CLT Sampling Approximation

Another approximation that we can use instead of full MC sampling is the conditional application of the Central Limit Theorem (CLT) which is valid if the

number of counterparties is large and the contribution of each counterparty is relatively small (granularity or "smallness" condition). Fewer scenarios are required for this approximation.

Under the CLT approximation, conditional losses (for each systemic sample l) are approximately normally distributed $\mathcal{N}(\mu_l(x), \sigma_l^2(x))$, where $\mu_l(x) = \mathbb{E}[\mathcal{L}(x) \mid Y = y_l]$ is the conditional mean of total portfolio loss and $\sigma_l(x)^2 = \mathrm{Var}[\mathcal{L}(x) \mid Y = y_l]$ is its conditional variance.

To compute $\mu_l(x)$ and $\sigma_l^2(x)$, first note that the conditional probability of a counterparty j being in credit state c, given that $Y = y_l$, is

$$\begin{aligned}
&\mathbb{P}(\text{CP } j \text{ is in credit state } c \mid Y = y_l) \\
&= \mathbb{P}(B_{c-1}^j \le \beta^j Y_{n(j)} + \sqrt{1-(\beta^j)^2} Z_j < B_c^j \mid Y = y_l) \\
&= \mathbb{P}\left(\frac{B_{c-1}^j - \beta^j y_l}{\sqrt{1-(\beta^j)^2}} \le Z_j < \frac{B_c^j - \beta^j y_l}{\sqrt{1-(\beta^j)^2}} \right) \\
&= \Phi\left(\frac{B_c^j - \beta^j y_l}{\sqrt{1-(\beta^j)^2}} \right) - \Phi\left(\frac{B_{c-1}^j - \beta^j y_l}{\sqrt{1-(\beta^j)^2}} \right) \equiv p_{c,l}^j,
\end{aligned}$$

where Φ is the standard normal cdf.

Under a given systemic scenario l, the conditional mean and variance of the loss due to the initial position with counterparty j are given by

$$\begin{aligned}
\mu_l^j &= \sum_{c \ge 0} \ell_c^j p_{c,l}^j, \\
(\sigma_l^j)^2 &= \sum_{c \ge 0} (\ell_c^j)^2 p_{c,l}^j - (\mu_l^j)^2,
\end{aligned}$$

The conditional mean and variance of loss from the i-th group are

$$\mu_{i,l} = \sum_{j \in G_i} \mu_l^j, \qquad \sigma_{i,l}^2 = \sum_{j \in G_i} (\sigma_l^j)^2.$$

The portfolio's conditional mean and variance of loss are

$$\mu_l(x) = \sum_{i=1}^{N_G} \mu_{i,l} x_i, \tag{6.3.9}$$

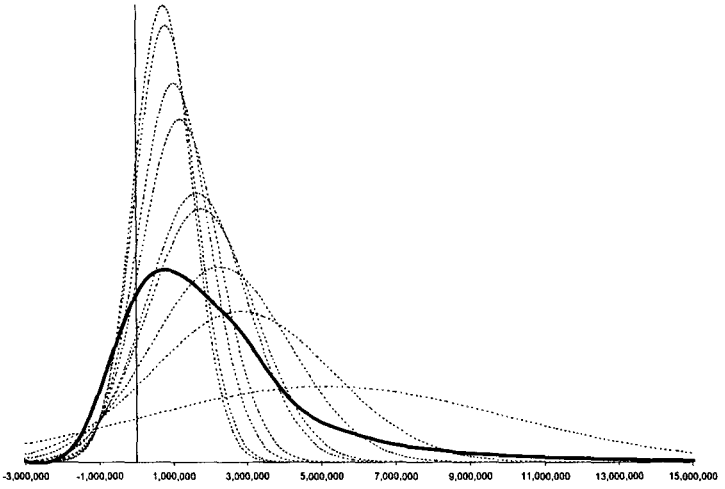


Figure 6.4: CLT Approximation of an Unconditional Loss Distribution

and

$$\sigma_l(x)^2 = \sum_{i=1}^{N_G} (\sigma_{i,l})^2 x_i^2. \tag{6.3.10}$$

The CLT approximation of the conditional portfolio loss distribution, is

$$\mathbb{P}_l(\mathcal{L}(x) \leq \ell) \approx \Phi\left(\frac{\ell - \mu_l(x)}{\sigma_l(x)}\right), \tag{6.3.11}$$

where $\mathbb{P}_l$ denotes the conditional probability measure with $\mathbb{P}_l(\mathcal{L}(x) \leq \ell) = \mathbb{P}(\mathcal{L}(x) \leq \ell \mid Y = y_l)$.

The approximation to the unconditional distribution is then a mixture of normal distributions and its cdf is equal to

$$\hat{F}^{\mathrm{CLT}}(\ell; x) = \frac{1}{M} \sum_{l=1}^{M} \Phi\left(\frac{\ell - \mu_l(x)}{\sigma_l(x)}\right). \tag{6.3.12}$$

The resulting approximation for the unconditional loss distribution is illustrated in Figure 6.4. The dashed curves are the conditional normal distributions of portfolio losses for $M = 9$ scenarios and the solid line is the unconditional loss distribution.

6.3.3.3 LLN Sampling Approximation

For a portfolio with a very large number of small counterparties we can use the Law of Large Numbers (LLN) to estimate conditional portfolio losses. In this case we assume that all specific risk is diversified away so that the portfolio loss is the sum of expected losses. As a result, as $N_{\text{CP}} \to \infty$ the conditional loss distribution is dominated by the mean loss over that scenario. As with the CLT conditional approximation, fewer scenarios are required for this approximation than for the MC approximation.

The LLN-approximation[3] of the conditional distribution of losses, is completely described by its mean, given by equation (6.3.9). The unconditional loss distribution is approximated by

$$\hat{F}^{\text{LLN}}(\ell; x) = \frac{1}{M} \sum_{l=1}^{M} \mathbf{1}\{\mu_l(x) \leq \ell\}. \tag{6.3.13}$$

6.3.4 Risk Measures

Our primary interest is in quantile-based risk measures such as value-at-risk and expected shortfall. For a detailed review of these risk measures we refer the reader to Section 6.1. In addition to minimizing VaR and ES directly, we also consider minimizing variance and the second moment of losses, as alternative, indirect ways to reduce quantile-based risk. Our goal is to compare the performance of credit-risk optimization when we optimize over different risk measures and evaluate VaR and ES for the resulting optimal portfolios.

The estimated vector of unconditional mean losses is $\mathbb{E}[\mathcal{L}(x)]$ and the unconditional variance of losses is $\text{Var}[\mathcal{L}(x)]$. Mean-variance (moment-based) optimization problems minimize variance or a combination of mean and variance.

[3]The usual interpretation of the LLN, would be that, under $\mathbb{P}_l$, $\lim_{N_{\text{CP}}\to\infty} \mathcal{L}(x)/N_{\text{CP}} = \lim_{N_{\text{CP}}\to\infty} \mu_l(x)/N_{\text{CP}}$ (assuming the latter limit exists), with $\mu_l(x)$ given by (6.3.9). We are using the term, LLN-approximation, to simply mean that under $\mathbb{P}_l$, $\mathcal{L}(x) \approx \mu_l(x)$.

Another risk measure that can be used instead of variance is the *second moment*, which is $\mathbb{E}[\mathcal{L}(x)^2] = \mathrm{Var}[\mathcal{L}(x)] + \mathbb{E}[\mathcal{L}(x)]^2$.

Value-at-risk is the maximum loss of a portfolio over a given time period and at a given level of probability. The value-at-risk function $\ell_\alpha(x)$ is the α-quantile of the loss distribution and is given by

$$\ell_\alpha(x) = \min\{\ell \in \mathbb{R} \,:\, \mathbb{P}(\mathcal{L}(x) \leq \ell) \geq \alpha\}.$$

Then for our structural model we have

$$\alpha = \mathbb{P}(\mathcal{L}(x) \leq \ell_\alpha(x)) = \int \mathbb{P}_y(\mathcal{L}(x) \leq \ell_\alpha(x))\, \mathrm{d}\varphi(y) \approx \frac{1}{M} \sum_{l=1}^{M} \mathbb{P}_l(\mathcal{L}(x) \leq \ell_\alpha(x)), \tag{6.3.14}$$

where φ is the distribution of the vector of credit drivers.

The expected shortfall $\mathrm{ES}_\alpha(x)$ is defined as the expected loss exceeding VaR and it can be written as

$$\mathrm{ES}_\alpha(x) = \frac{1}{1-\alpha} \mathbb{E}\left[\mathcal{L}(x) \cdot \mathbf{1}\{\mathcal{L}(x) \geq \ell_\alpha(x)\}\right]. \tag{6.3.15}$$

For our structural model we have

$$\mathbb{E}\left[\mathcal{L}(x) \cdot \mathbf{1}\{\mathcal{L}(x) \geq \ell_\alpha(x)\}\right] \approx \frac{1}{M} \sum_{l=1}^{M} \mathbb{E}_l\left[\mathcal{L}(x) \cdot \mathbf{1}\{\mathcal{L}(x) \geq \ell_\alpha(x)\}\right],$$

where $\mathbb{E}_l$ denotes the conditional expectation operator $\mathbb{E}[\,\cdot \mid Y = y_l]$.

The goal of minimizing the risk measures discussed above, leads us to the general formulation of the optimization problem. All the optimization formulations we are going to discuss in Section 6.3.5 will have the following general form:

$$\begin{array}{ll} \min\limits_{x} & g\left[\mathcal{L}(x)\right] \\ \text{s.t.} & x \in \Omega, \end{array} \tag{6.3.16}$$

where $g\left[\mathcal{L}(x)\right]$ is the risk measure ($\mathrm{Var}[\mathcal{L}(x)]$, $\mathbb{E}[\mathcal{L}(x)^2]$, $\ell_\alpha(x)$ or $\mathrm{ES}_\alpha(x)$) and Ω denotes the feasible region, defined by a set of linear constraints.

6.3.5 Credit Risk Optimization Formulations

In this section we formulate the optimization problems for minimizing risk measures described in Section 6.3.4 within the framework introduced in Sections 6.3.2 and 6.3.3. These two dimensions, the risk measure and the approximation of the loss distribution, define the taxonomy of the optimization problems that we consider.

For ease of presentation, from here onwards we suppress the subscript G on N_G, the number of groups, and write simply N.

6.3.5.1 Moment-Based Formulations

Minimizing variance or the second moment of the loss distribution does not minimize the quantile-based risk measures (VaR and ES), unless the distribution of losses is normal. So, it is expected that the moment-based formulations may not perform well for all quantiles.

For the moment-based formulations, we require the vector, μ, of unconditional expected credit losses of the groups, and the unconditional variance-covariance matrix, Q, of the groups' credit losses. The i-th component of μ, indexed by the groups, is the sum over the counterparties in the i-th group, of the counterparties' mean losses:

$$\mu_i = \sum_{j \in G_i} \sum_{c=0}^{C-1} \ell_c^j p_c^j.$$

The (i_1, i_2) component of Q is the sum of the covariances of the counterparty losses, with the counterparties ranging over the i_1-th and i_2-th groups. MC approximation or semi-analytically, using the conditional independence of counterparty losses combined with a systemic MC simulation for unconditioning.

The variance $\mathrm{Var}[\mathcal{L}(x)]$ minimization problem is:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^N} & x^T Q x \\ \text{s.t.} & x \in \Omega. \end{array}$$

The second moment $\mathbb{E}[\mathcal{L}(x)^2]$ minimization problem is:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^N} & x^T[Q+\mu\mu^T]x \\ \text{s.t.} & x\in\Omega. \end{array}$$

6.3.5.2 ES and VaR Minimization with MC-Sampling Approximation

Optimization Formulation for ES Minimization

It was shown in [121] and [122] that minimization of expected shortfall $\mathrm{ES}_\alpha(x)$ can be reduced to minimizing the function

$$\ell+\frac{1}{1-\alpha}\mathbb{E}([\mathcal{L}(x)-\ell]^+), \tag{6.3.17}$$

where $a^+=\max(0,a)$.

The function (6.3.17) is convex in ℓ; it is also convex in x if the function of losses, $\mathcal{L}(x)$, is convex in x.

Having MK scenarios with corresponding probabilities $1/MK$ of occurring, we can approximate the expectation in (6.3.17):

$$\mathbb{E}([\mathcal{L}(x)-\ell]^+)\approx\frac{1}{MK}\sum_{l,k}[L_{lk}(x)-\ell]^+. \tag{6.3.18}$$

The ES optimization problem becomes:

$$\begin{array}{ll} \min\limits_{x,\ell} & \ell+\dfrac{1}{(1-\alpha)}\dfrac{1}{MK}\displaystyle\sum_{l,k}^{MK}[L_{lk}(x)-\ell]^+ \\ \text{s.t.} & x\in\Omega. \end{array}$$

The problem is convex as the loss functions $L_{lk}(x)$ are linear and the set Ω is defined by linear equalities and inequalities. The ES minimization problem can be reduced to the linear problem:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^N,\,u\in\mathbb{R}^{MK},\,\ell\in\mathbb{R}} & \ell+\dfrac{1}{(1-\alpha)}\dfrac{1}{MK}\displaystyle\sum_{l,k}u_{lk}, \\ \text{s.t.} & u_{lk}\geq L_{lk}(x)-\ell,\ u_{lk}\geq 0,\ l=1,\ldots,M,\quad k=1,\ldots,K \\ & x\in\Omega, \end{array} \tag{6.3.19}$$

where $\{u_{lk}\}$ are auxiliary variables.

Note that to optimize the expectation in (6.3.17), we can use a number of approaches. Instead of formulating the problem as a large-scale linear optimization problem described in this section, it can be solved as a non-smooth, nonlinear optimization problem. In general, using nonsmooth optimization techniques is prohibitively slow. Instead, we can use the smoothing technique in [2] to solve the problem with a smoothed ES function.

Sequential Algorithm for VaR Minimization

For the MC-sampling formulation, directly minimizing VaR requires integer programming. We use a heuristic technique, developed in [87], that is less computationally intensive. The algorithm minimizes VaR by solving a sequence of ES minimization problems while progressively fixing scenarios in the tail of the loss distribution.

In our setting, the algorithm is described by Algorithm 6.1.

6.3.5.3 VaR and ES Minimization Based on CLT Approximation

Applying the CLT approximation to the conditional loss distribution, from (6.3.11) we obtain

$$\mathbb{P}_l(\mathcal{L}(x) \leq \ell_\alpha(x)) \approx \Phi\left(\frac{\ell_\alpha(x) - \mu_l(x)}{\sigma_l(x)}\right).$$

From (6.3.14), we derive:

$$\frac{1}{M}\sum_{l=1}^{M} \Phi\left(\frac{\ell_\alpha(x) - \mu_l(x)}{\sigma_l(x)}\right) = \alpha. \tag{6.3.20}$$

Equation (6.3.20) determines the value of the objective function, $\ell_\alpha(x)$, implicitly as function of the decision vector x. For fixed x, the VaR value $\ell_\alpha(x)$ is the solution of equation (6.3.20) which must be solved numerically, for instance using bisection or Newton-type methods.

This leads to the following formulation of the VaR optimization problem

Algorithm 6.1: Sequential Algorithm for VaR Minimization.

Step 0. Initialization

1. Set $\alpha_0 = \alpha$, $m = 0$, $H_0 = \{(l,k) : l = 1,\dots,M,\ k = 1,\dots,K\}$.
2. Assign a value to the parameter, ε, for discarding scenarios; $0 < \varepsilon < 1$.

Step 1. Optimization subproblem

1. Minimize α_m-ES

$$\begin{array}{lll} \min\limits_{x,u,\ell,\gamma} & \ell + \nu_m \frac{1}{MK} \sum\limits_{(l,k)\in H_m} u_{lk} & \\ \text{s.t.} & L_{lk}(x) \le \ell + u_{lk},\ u_{lk} \ge 0 & (l,k) \in H_m, \\ & L_{lk}(x) \le \gamma & (l,k) \in H_m, \\ & L_{lk}(x) \ge \gamma & (l,k) \notin H_m, \\ & x \in \Omega, & \end{array}$$

   where $\nu_m = 1/((1-\alpha_m))$. Denote the optimal solution of this problem by x^*_m.
2. Denote the order statistics of the losses $L_{lk}(x^*_m)$, $l = 1,\dots,M,\ k = 1,\dots,K$ by L_n, $n = 1,\dots,MK$. Also, denote the sorting order by writing $n(l,k) = n$ if (l,k) is the n-th index in the sorting order.

Step 2. Estimating VaR
Calculate VaR estimate $\ell_m = L_{n_\alpha}$, where $n_\alpha = \min\{n : n/MK \ge \alpha\}$.

Step 3. Stopping and re-initialization

1. $m = m + 1$.
2. $b_m = \alpha + (1-\alpha)(1-\varepsilon)^m$ and $\alpha_m = \alpha/b_m$.
3. $H_m = \{(l,k) \in H_{m-1} : n(l,k)/MK \le b_m\}$.
4. If $H_m = H_{m-1}$ then stop the algorithm and return the estimate of the VaR-optimal portfolio x^*_m and VaR estimate ℓ_m, otherwise go to Step 1.

with the conditional normal approximation:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^N} & \ell_\alpha(x) \\ \text{s.t.} & \frac{1}{M}\sum_{l=1}^{M}\Phi\left(\frac{\ell_\alpha(x)-\mu_l(x)}{\sigma_l(x)}\right)=\alpha, \\ & x\in\Omega. \end{array} \tag{6.3.21}$$

Turning to ES, under scenario l, $\mathcal{L}(x) \sim \mathcal{N}(\mu_l(x), \sigma_l^2(x))$. Now, the ES for a general $\mathcal{N}(\mu, \sigma^2)$-distributed random variable, X, is well known to be

$$\frac{1}{1-\alpha}\left[\mu\overline{\Phi}\left(\frac{x_\alpha-\mu}{\sigma}\right)+\sigma\phi\left(\frac{x_\alpha-\mu}{\sigma}\right)\right],$$

where x_α is the α-quantile of X, ϕ denotes the standard normal pdf, and $\overline{\Phi} = 1-\Phi$.

Therefore, the expected shortfall $\mathrm{ES}_\alpha(x)$ at the quantile level α is

$$\mathrm{ES}_\alpha(x) = \frac{1}{1-\alpha}\frac{1}{M}\sum_{l=1}^{M}\left[\mu_l(x)\overline{\Phi}\left(\frac{\ell_\alpha(x)-\mu_l(x)}{\sigma_l(x)}\right)+\sigma_l(x)\phi\left(\frac{\ell_\alpha(x)-\mu_l(x)}{\sigma_l(x)}\right)\right]. \tag{6.3.22}$$

The resulting ES optimization problem with CLT approximation is:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^N} & \mathrm{ES}_\alpha(x) \\ \text{s.t.} & \frac{1}{M}\sum\limits_{l=1}^{M}\Phi\left(\frac{\ell_\alpha(x)-\mu_l(x)}{\sigma_l(x)}\right)=\alpha, \\ & x\in\Omega, \end{array} \tag{6.3.23}$$

where $\mathrm{ES}_\alpha(x)$ is defined by equation (6.3.22).

Numerical solutions for problems (6.3.21) and (6.3.23) can be obtained using standard nonlinear optimization techniques.

6.3.5.4 ES and VaR Minimization with LLN Approximation

The LO problem for ES minimization was defined by (6.3.19). The only difference between the MC and LLN approximation formulations is the loss function $\mathcal{L}(x)$. Using the loss function $\mathcal{L}_l(x)$ for the LLN approximation, we get the following

ES optimization problem:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^N,\ u\in\mathbb{R}^M,\ \ell\in\mathbb{R}} & \ell + \dfrac{1}{(1-\alpha)}\dfrac{1}{M}\displaystyle\sum_{l=1}^{M} u_l \\ \text{s.t.} & u_l \geq \mu_l(x) - \ell,\ u_l \geq 0,\ l = 1,\ldots,M \\ & x \in \Omega, \end{array} \tag{6.3.24}$$

where $\mu_l(x)$ is the mean loss vector from LLN scenarios ($l = 1,\ldots,M$), computed from (6.3.9).

The algorithm described in Section 6.3.5.2 is suitable for VaR minimization with the conditional LLN approximation. The modification of it for the LLN approximation is straightforward.

6.3.6 Taxonomy of Optimization Problems and Data Requirements

The taxonomy of optimization problems, that we have described in the previous subsections, is presented in Table 6.2. The empty cells that appear in Table 6.2 are due to two reasons. First, variance optimization under conditional independence framework is equivalent to the unconditional variance optimization formulation. Second, for unconditional formulation VaR and ES minimization is equivalent to variance minimization.

According to our knowledge, formulations for VaR and ES optimization problems within a conditional independence framework appear in [77] and in this thesis for the first time. Nonlinear optimization formulations for CLT approximation are novel, that is especially important for VaR optimization as finding exact solutions to that class of problems requires utilizing mixed-integer optimization techniques.

Data requirements for optimization by each formulation can be summarized as follows:

- MC Sampling - M systemic scenarios, K specific scenarios for each systemic, N groups: KMN data points;

Table 6.2: The Taxonomy of Credit Risk Optimization Problems.

Optimization Formulation \ Risk Measure	VaR	Expected Shortfall	Variance Second Moment
CLT Approximation	***Non-convex nonlinear problem*** Direct VaR minimization with CLT sampling (Section 6.3.5.3)	***Convex nonlinear problem*** Direct expected shortfall minimization with CLT sampling (Section 6.3.5.3)	
LLN Approximation	***Linear problem (heuristic)*** Successive expected shortfall minimization heuristics with LLN sampling (Section 6.3.5.4)	***Linear problem*** Direct expected shortfall minimization with LLN sampling (Section 6.3.5.4)	
Monte Carlo Sampling	***Linear problem (heuristic)*** Successive expected shortfall minimization heuristics with Monte-Carlo sampling (Section 6.3.5.2)	***Linear problem*** Direct expected shortfall minimization with Monte-Carlo sampling (Section 6.3.5.2)	
Unconditional			***Convex quadratic problem*** Vector of mean losses and variance-covariance matrix of losses are computed from Monte-Carlo sampling (Section 6.3.5.1)

- CLT Approximation - M systemic scenarios, 1 mean and 1 variance for each systemic, N groups: $2MN$ data points;
- LLN Approximation - M systemic scenarios, 1 mean for each systemic, N groups: MN data points.

It is not surprising that the LLN formulation uses the least amount of data points for the optimization, as it is the most restrictive formulation, relying on the portfolio being very large and highly granular. If $K > 2$, the MC-sampling formulation requires the largest number of data points for optimization.

Computational results for all formulations in Table 6.2 are presented in Appendix A.

Chapter 7

Case Studies in Finance

In this chapter we present examples of multiobjective optimization problems from finance that can be formulated and solved via parametric optimization. Those examples and their corresponding parametric formulations are discussed in the following sections. We describe our computational results on both small size illustrative case studies and large scale real problems.

7.1 Portfolio Selection with Multiple Linear Objectives

Here, we discuss the multiobjective portfolio selection problem, where the objective functions are linear. Those models are rooted in the Capital Asset Pricing Model described in Section 6.1.1.2, where the risk measure of an asset or portfolio is given by its beta coefficient. CAPM is the equilibrium version of mean-variance theory. Due to measuring risk in terms of the beta coefficients, the objective function in the risk minimization problem is linear in portfolio weights. In [160] a decision tool for the selection of stock portfolios based on multiobjective LO is described. Linear objective functions of the problem are the return, price earnings ratio, volume of transactions, dividend yield, increase in profits and risk, which is expressed as the linear function of betas. The authors apply portfolio selection to a set of fifty-two stocks from the Athens Stock Exchange. We are go-

ing to briefly describe their model including objective functions and constraints and compute the Pareto front for three out of six objectives considered in [160]. Readers interested in full details of the formulation and data for the model may consult [160].

The decision variables in portfolio selection problems are the portfolio weights x_i, $i = 1, \ldots, N$, where N is the total number of assets available for investment. Portfolio weights define a proportion of total wealth (or total budget) invested in the corresponding stock. As a matter of convenience, sum of portfolio weights is normalized to one $\sum_{i=1}^{N} x_i = 1$. Denoting by r_i the expected market return of an asset i, allows us to compute the portfolio market return as $r_P = \sum_{i=1}^{N} r_i x_i = r^T x$.

The beta coefficient β is a relative measure of systematic (non-diversifiable) risk, it reflects the tendency of an asset to move with the market. As beta measures correlation with the market portfolio, it is calculated as $\beta_i = \frac{\mathrm{Cov}(r_i, r_M)}{\mathrm{Var}(r_M)}$, where r_i is the asset i return and r_M is the return of the market portfolio. If $\beta_i < 1$ then asset i has less systematic risk than the overall market and the opposite holds for $\beta_i > 1$. As a result, portfolio risk minimization can be expressed as the linear function of asset weights, namely $\{\min_x \beta^T x\}$.

Among the other six objectives that are considered in [160] is maximizing return $\{\max_x r^T x\}$ and minimizing Price Earnings Ratio (P/E) $\{\min_x d^T x\}$. The Price Earnings Ratio d_i for each stock is computed as share price in the stock market at time period t divided by earnings per share at period $t-1$. We could have computed the Pareto efficient surface for more than three objectives here, but we restrict our attention to only those three due to well known difficulties with visualizing surfaces in more than 3 dimensions. Denoting the three objectives as $f_1 = -r^T x$, $f_2 = \beta^T x$ and $f_3 = d^T x$, we obtain the following parametric optimization problem:

$$\begin{array}{ll} \min & -r^T x + \lambda_1 \beta^T x + \lambda_2 d^T x \\ \text{s.t.} & x \in \Omega, \end{array} \tag{7.1.1}$$

where Ω in [160] is the set of linear constraints that includes no-short-sales re-

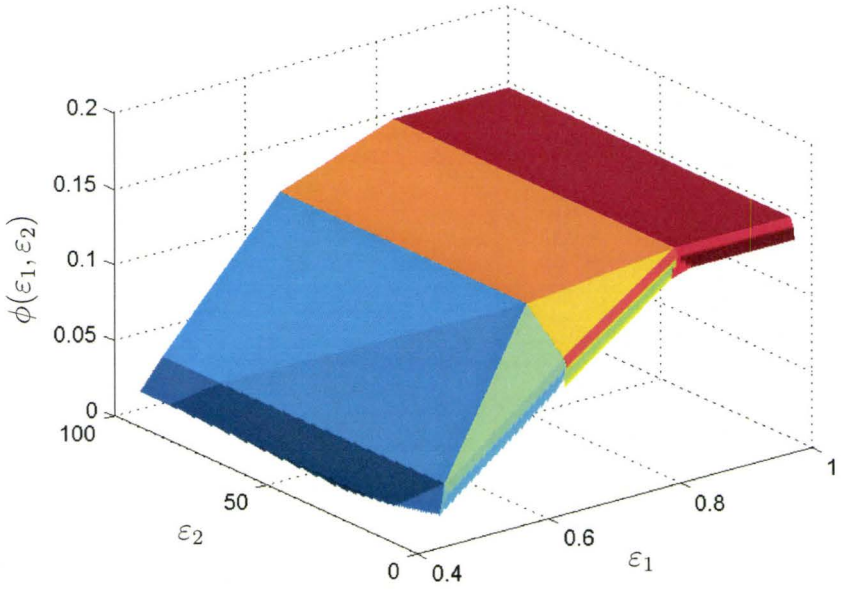


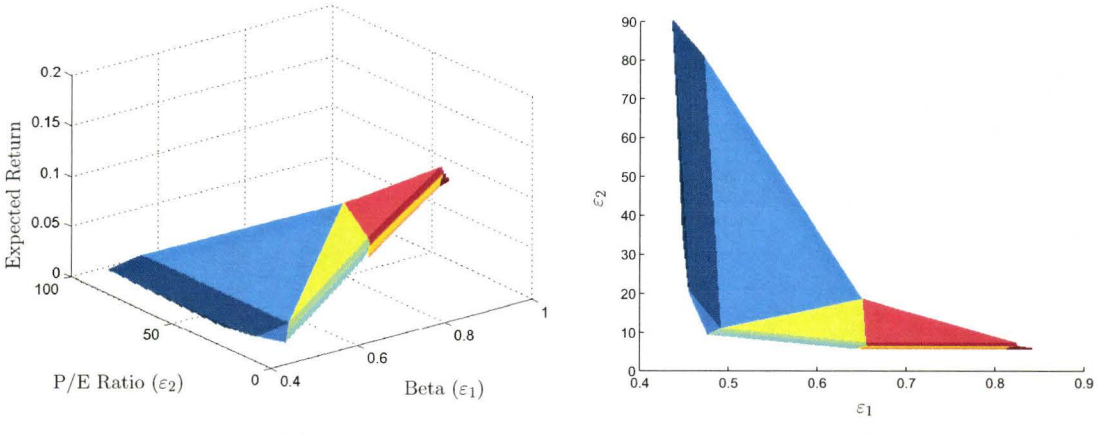
Figure 7.1: The Optimal Value Function for the Parametric Linear Portfolio Optimization Problem.

striction $x \geq 0$; upper limits for the capital allocations $x_i \leq u_i$, $i = 1, \ldots, 52$; specific preferences for some stocks of the form $x_j \geq l_i$; and the constraints on betas of the form that portion y of the capital will be allocated to stocks with $\beta \in \{\beta_1, \beta_2\}$ that are expressed as $\sum_{i \subseteq I} x_i = y$. Note that maximizing $r^T x$ is equivalent to minimizing $-r^T x$.

The parametric optimization problem that follows from the ε-constrained multiobjective formulation is the following:

$$\begin{array}{ll} \min\limits_{x,t} & -r^T x \\ \text{s.t} & \beta^T x + t_1 = \varepsilon_1 \\ & d^T x + t_2 = \varepsilon_2 \\ & \sum_i x_i = 1 \\ & \sum_{i \in I} x_i = 0.2 \\ & x \geq 0,\ t \geq 0, \end{array} \tag{7.1.2}$$

where t_1, t_2 are the slack variables used to convert the linear inequality constrains



(a) (b)

Figure 7.2: The Pareto Front for the Multiobjective Linear Portfolio Optimization Problem (a) and the Invariancy Regions Corresponding to It (b).

into equality constraints and $\varepsilon = (\varepsilon_1, \varepsilon_2)^T$ is the vector of parameters. We have used a subset of the constraints $x \in \Omega$ from [160] for the ease of exposition and included the no short-sales constraint $x \geq 0$ and the constraint $\sum_{i \in I} x_i = 0.2$ stating that 20% of capital is allocated to stocks with a beta coefficient less than 0.5. Formulation (7.1.2) is parametric LO problem with two parameters in the right-hand-side of the constraints.

The optimal value function for problem (7.1.2) is shown in Figure 7.1. We can use the optimal partition for the variables t_1 and t_2 to determine the Pareto-efficient surface. For the invariancy regions corresponding to Pareto-efficient solutions, $t_1 \in \mathcal{N}$ and $t_2 \in \mathcal{N}$, meaning that those variables belong to the subset $\mathcal{N}$ of the optimal partition. The invariancy regions corresponding to the Pareto efficient solutions are shown in Figure 7.2(b) and the Pareto front is depicted in Figure 7.2(a). The Pareto front is a piecewise linear function. The knowledge of invariancy intervals and optimal value function on those intervals gives us the structure of the Pareto front.

7.2 Mean-Variance Optimization with Market Risk

The Markowitz mean-variance model is commonly used in practice in the presence of market risk. From an optimization perspective, minimizing variance requires solving a QO problem. Denoting a vector of expected market returns by r as before and a variance-covariance matrix of returns by Q, the mean-variance portfolio optimization problem is formulated as a QO problem where the objectives are to maximize the expected portfolio return $\{\max_x r^T x\}$ and to minimize variance $\{\min_x x^T Q x\}$. The multiobjective optimization problem can be formulated as the weighted sum problem

$$\begin{array}{ll} \min_x & \lambda r^T x + \frac{1}{2} x^T Q x \\ \text{s.t} & x \in \Omega, \end{array} \tag{7.2.1}$$

or as the ε-constrained problem

$$\begin{array}{ll} \min_x & \frac{1}{2} x^T Q x \\ \text{s.t.} & -r^T x \leq \varepsilon, \\ & x \in \Omega, \end{array} \tag{7.2.2}$$

where Ω is the set of linear constraints on portfolio weights.

We use a small portfolio optimization problem to illustrate the multiobjective optimization methodology. Problem data is presented in Table 7.1 that shows the vector of expected market returns r and the variance-covariance matrix Q for the portfolio of 5 securities. We put non-negativity bounds $x \geq 0$ on the weights disallowing short-sales and add a constraint that makes the sum of the weights equal to 1. Two conflicting objectives are considered:

1) minimize the variance of returns;
2) maximize expected market return.

Thus, the multiobjective portfolio optimization problem becomes:

$$\begin{array}{ll} \min_x & f_1(x) = -r^T x,\ f_2(x) = \frac{1}{2} x^T Q x \\ \text{s.t} & \sum_i x_i = 1, \\ & x_i \geq 0\ \forall i. \end{array} \tag{7.2.3}$$

Table 7.1: The Expected Return Vector r and the Return Covariance Matrix Q for the Mean-Variance Optimization Problem with Market Risk.

Security	Expected Return
1	0.096268
2	0.069400
3	0.088758
4	0.111474
5	0.043281

Security	Variance-Covariance Matrix				
1	0.008989	0.002727	0.003838	0.007222	0.003944
2	0.002727	0.004982	0.002150	0.002191	0.003018
3	0.003838	0.002150	0.009153	0.005645	0.004704
4	0.007222	0.002191	0.005645	0.016891	0.004047
5	0.003944	0.003018	0.004704	0.004047	0.005156

We solve two formulations of problem (7.2.3). The first one is the parametric problem corresponding to the ε-constraint multiobjective formulation:

$$\begin{array}{ll} \min\limits_{x,t} & \frac{1}{2}x^TQx \\ \text{s.t} & -r^Tx + t = \varepsilon \\ & \sum_i x_i = 1, \\ & x \geq 0,\ t \geq 0, \end{array} \tag{7.2.4}$$

where t is the slack variable used to convert the linear inequality constraint into equality constraint and ε is the parameter. The second formulation is the parametric QO problem corresponding to the weighting method:

$$\begin{array}{ll} \min\limits_{x} & -\lambda r^Tx + \frac{1}{2}x^TQx \\ \text{s.t} & \sum_i x_i = 1, \\ & x \geq 0, \end{array} \tag{7.2.5}$$

where λ is the parameter.

The output of our parametric solver for the ε-constrained formulation (7.2.4) is shown in Table 7.2, while Table 7.3 displays the output for the formulation (7.2.5). While both formulations produce the same efficient frontier, see Figure 7.4, we may notice that formulation (7.2.5) produces two optimal partitions $\mathcal{B}_1 = 4$, $\mathcal{N}_1 = \{2,3,5\}$, $\mathcal{T}_1 = 1$ and $\mathcal{B}_2 = 4$, $\mathcal{N}_1 = \{1,2,3,5\}$, $\mathcal{T}_1 = \varnothing$ for the right-most point on the frontier, while formulation (7.2.4) produces only one optimal partition $\mathcal{B}_1 = 4$, $\mathcal{N}_1 = \{2,3,5\}$, $\mathcal{T}_1 = 1$ at the same point. Note that index 6 in the optimal partitions in Table 7.2 is for the slack variable t.

The optimal value function $\phi(\varepsilon)$ for formulation (7.2.4) is shown in Figure 7.3. The Markowitz efficient frontier in financial terminology or the Pareto

Table 7.2: The Output of the Parametric Solver for the Mean-Variance Problem with Market Risk, ε-Constrained QO Formulation.

type	ε_ℓ	ε_u	$\mathcal{B}$	$\mathcal{N}$	$\mathcal{T}$	$\phi(\varepsilon)$
inv. interv	+0.00000	+0.06865	1 2 3 4 5 6			$0.00000\varepsilon^2$-0.00000ε+0.00194
tr. point	+0.06865	+0.06865	1 2 3 4 5		6	0.00194
inv. interv	+0.06865	+0.07974	1 2 3 4 5	6		$0.51346\varepsilon^2$-0.07049ε+0.00435
tr. point	+0.07974	+0.07974	1 2 3 4	6	5	0.00200
inv. interv	+0.07974	+0.09891	1 2 3 4	5 6		$4.24940\varepsilon^2$-0.66632ε+0.02811
tr. point	+0.09891	+0.09891	1 3 4	5 6	2	0.00378
inv. interv	+0.09891	+0.10620	1 3 4	2 5 6		$13.77287\varepsilon^2$-2.55033ε+0.12129
tr. point	+0.10620	+0.10620	1 4	2 5 6	3	0.00578
inv. interv	+0.10620	+0.11147	1 4	2 3 5 6		$24.72911\varepsilon^2$-4.87744ε+0.24486
tr. point	+0.11147	+0.11147	4	2 3 5 6	1	0.00845

Table 7.3: The Output of the Parametric Solver for the Mean-Variance Problem with Market Risk, Weighted Sum QO Formulation.

type	λ_ℓ	λ_u	$\mathcal{B}$	$\mathcal{N}$	$\mathcal{T}$	$\phi(\lambda)$
inv. interv	+0.00000	+0.01140	1 2 3 4 5			$-0.48690\lambda^2$-0.06865λ+0.00194
tr. point	+0.01140	+0.01140	1 2 3 4		5	0.00109
inv. interv	+0.01140	+0.17433	1 2 3 4	5		$-0.05883\lambda^2$-0.07840λ+0.00199
tr. point	+0.17433	+0.17433	1 3 4	5	2	-0.01346
inv. interv	+0.17433	+0.37504	1 3 4	2 5		$-0.01815\lambda^2$-0.09259λ+0.00323
tr. point	+0.37504	+0.37504	1 4	2 5	3	-0.03405
inv. interv	+0.37504	+0.63585	1 4	2 3 5		$-0.01011\lambda^2$-0.09862λ+0.00436
tr. point	+0.63585	+0.63585	4	2 3 5	1	-0.06244
inv. interv	+0.63585	+Inf	4	1 2 3 5		$0.00000\lambda^2$-0.11147λ+0.00845

front in multiobjective optimization terminology is depicted in Figure 7.4. The optimal value function is plotted in the variance – expected return coordinate system, while the efficient frontier is shown in the expected return – standard deviation coordinate system. Adjacent *corner portfolios* define a segment of the efficient frontier within which portfolios hold identical assets and the rate of change of asset weights in moving from one portfolio to another is constant [133]. Corner portfolios are identified by the Markowitz critical line algorithm [96]. Due to the uniqueness of the optimal solution for each parameter value in our example, corner portfolios on the efficient frontier in Figure 7.4 are the transition points of the parametric problem solution. In between this corner portfolios the frontier is piecewise quadratic. Figure 7.4 shows the efficient frontier in the mean-standard

deviation space in order to be consistent with the existing literature. Note that the efficient frontier is a piecewise quadratic function in the mean-variance space.

Portfolio composition is shown in Figure 7.5. As the covariance matrix Q for our problem is positive definite, we get strictly convex objective functions in our formulations, and consequently we get unique optimal solutions $x^*(\varepsilon)$, and $x^*(\lambda)$. It allows us to plot unique portfolio composition, namely dependance of asset weights x^* on the expected portfolio return ε, see formulation (7.2.4). Figure 7.5 displays corner portfolios and invariancy intervals. On each invariancy interval function $x^*(\varepsilon)$ is linear which is depicted on the lower part of Figure 7.5. On each invariancy interval the portfolio composition is constant, meaning that assets with zero weights stay zero for the portfolio on the whole interval, while weights for non-zero holdings may change on the interval.

7.3 Mean-Variance Optimization with Market Risk and Transaction Cost

A portfolio may incur transaction cost associated with each trading. In this section we extend the case study presented in Section 7.2 to allow for linear transaction cost. Denoting the linear transaction cost by ℓ_i, we add the third objective of minimizing the trading cost $\ell^T x$ of a portfolio to the mean-variance portfolio optimization problems (7.2.1)-(7.2.2).

The problem data is presented in Tables 7.4 and 7.5. Table 7.4 shows expected market returns per unit transaction cost for 8 securities, as well as their weights in the initial portfolio x_0.

We put non-negativity bounds $x \geq 0$ on the weights disallowing short-sales and optimize three objectives:

1) minimize the variance of returns;
2) maximize expected market return;
3) minimize transaction cost.

Moreover, we also need to add a constraint that makes the sum of the weights equal to one.

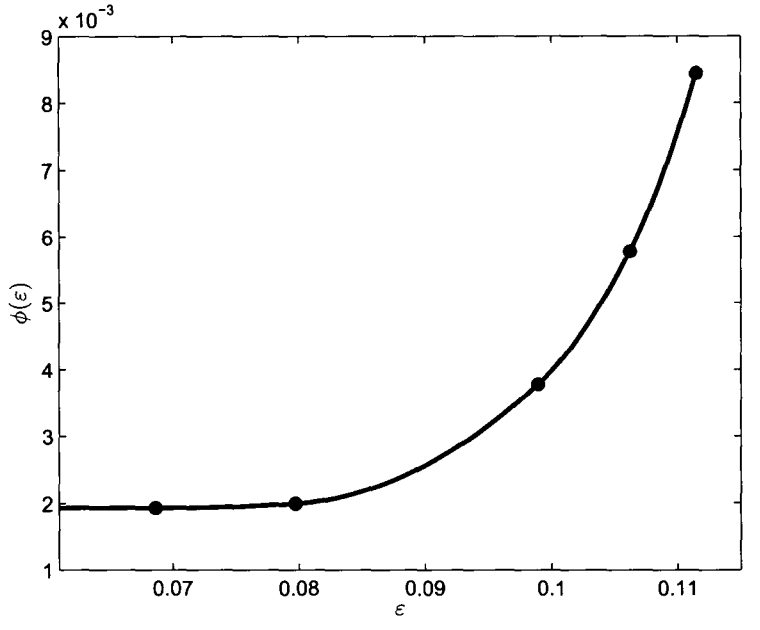


Figure 7.3: The Optimal Value Function for the Mean-Variance Problem with Market Risk.

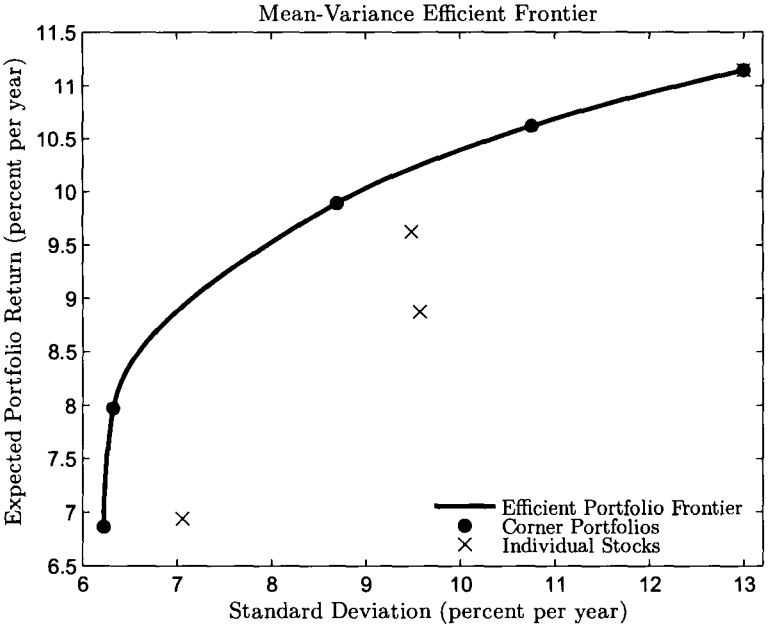


Figure 7.4: The Efficient Portfolio Frontier for the Mean-Variance Problem with Market Risk.

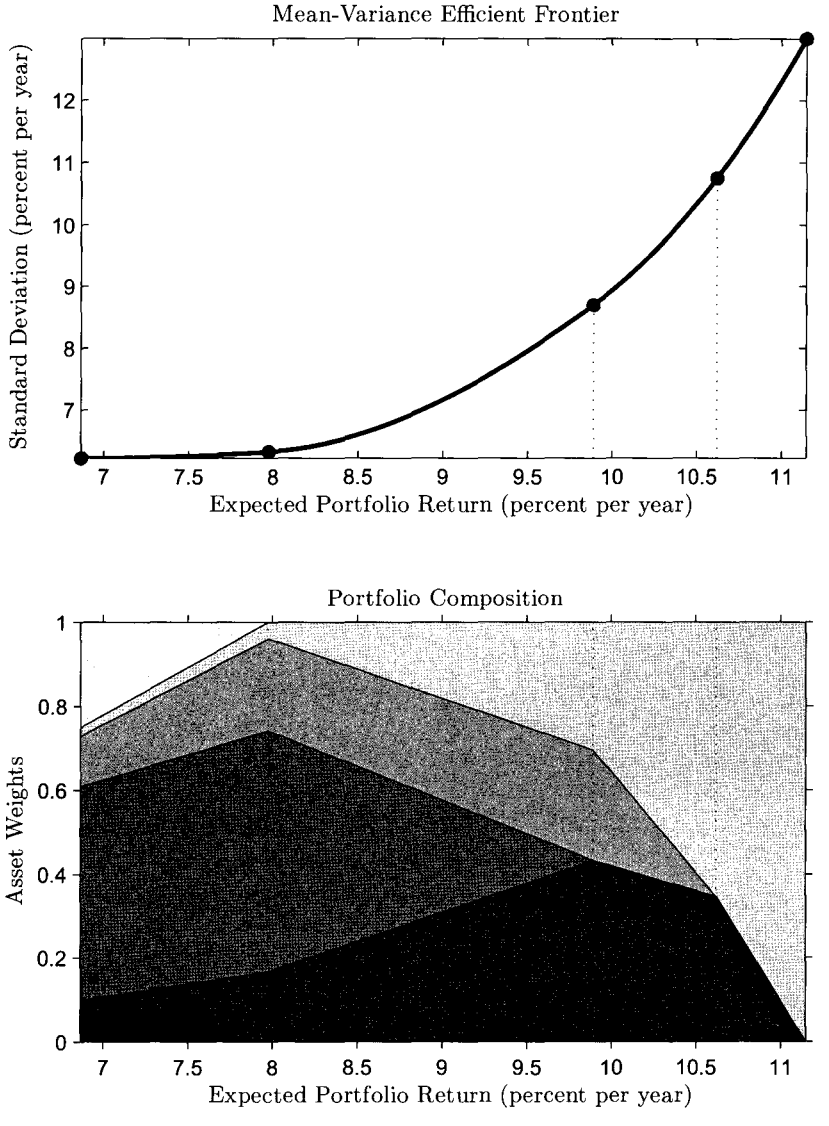


Figure 7.5: Portfolio Composition for the Mean-Variance Problem with Market Risk.

Table 7.4: Portfolio Data for Mean-Variance Optimization with Market Risk and Transaction Cost.

Security	x_0	$r = \mathbb{E}$(Market Return)	$\ell =$ (Transaction Cost)
1	0	0.095069	0.009830
2	0.44	0.091222	0.005527
3	0.18	0.140161	0.004001
4	0	0.050558	0.001988
5	0	0.079741	0.006252
6	0.18	0.054916	0.000099
7	0.13	0.119318	0.003759
8	0.07	0.115011	0.007334

Table 7.5: The Return Covariance Matrix Q for Mean-Variance Optimization with Market Risk and Transaction Cost.

Security	1	2	3	4	5	6	7	8
1	0.002812	0.002705	-0.001199	0.000745	-0.000064	0.001035	-0.000336	0.000178
2	0.002705	0.015664	-0.003000	0.001761	-0.002282	0.007129	0.000596	-0.003398
3	-0.001199	-0.003000	0.008842	-0.000155	0.003912	0.001424	0.001183	-0.001710
4	0.000745	0.001761	-0.000155	0.002824	0.001043	0.003874	0.000225	-0.001521
5	-0.000064	-0.002282	0.003912	0.001043	0.007213	-0.001946	0.001722	0.001199
6	0.001035	0.007129	0.001424	0.003874	-0.001946	0.013193	0.001925	-0.004506
7	-0.000336	0.000596	0.001183	0.000225	0.001722	0.001925	0.002303	-0.000213
8	0.000178	-0.003398	-0.001710	-0.001521	0.001199	-0.004506	-0.000213	0.006288

Thus, the multiobjective portfolio optimization problem looks like:

$$\begin{array}{ll} \min & f_1(x) = -r^T x,\ f_2(x) = \ell^T x,\ f_3(x) = \frac{1}{2} x^T Q x \\ \text{s.t} & \sum_i x_i = 1, \\ & x_i \geq 0\ \forall i. \end{array} \tag{7.3.1}$$

We solve problem (7.3.1) as a parametric problem corresponding to the ε-constraint multiobjective formulation:

$$\begin{array}{ll} \min\limits_{x,t} & \frac{1}{2} x^T Q x \\ \text{s.t} & -r^T x + t_1 = \varepsilon_1 \\ & \ell^T x + t_2 = \varepsilon_2 \\ & \sum_i x_i = 1, \\ & x \geq 0,\ t \geq 0, \end{array} \tag{7.3.2}$$

where t_1, t_2 are the slack variables used to convert the linear inequality constrains into equality constraints and $\varepsilon = (\varepsilon_1, \varepsilon_2)^T$ is the vector of parameters.

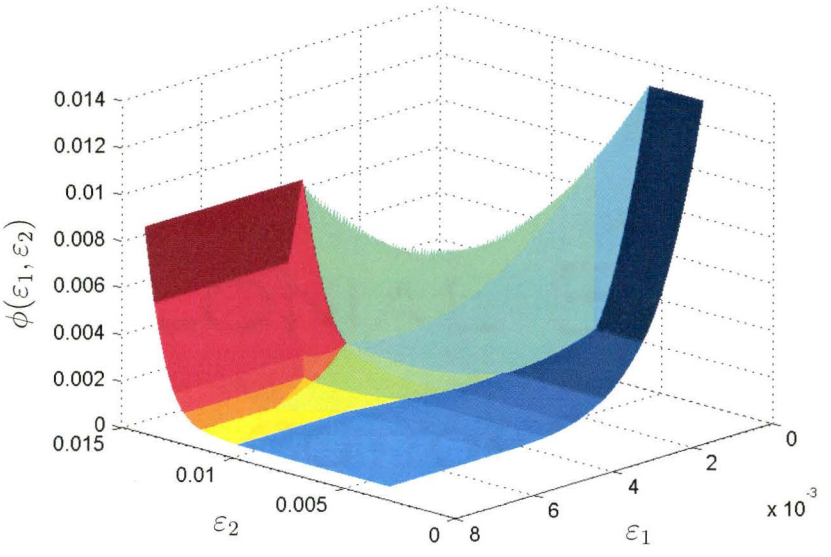


Figure 7.6: The Optimal Value Function for the Mean-Variance Portfolio Problem in the Presence of Transaction Cost.

The optimal value function for problem (7.3.2) is shown in Figure 7.6 and the corresponding invariancy regions – in Figure 7.7(a). We can utilize the optimal partition for the variables t_1 and t_2 to determine the Pareto efficient surface. For the invariancy regions corresponding to Pareto efficient solutions, $t_1 \neq \mathcal{B}$ and $t_2 \neq \mathcal{B}$, meaning that those variables do not belong to the subset $\mathcal{B}$ of the optimal partition. The invariancy regions corresponding to Pareto efficient solutions are shown in Figure 7.7(b) and the Pareto front is depicted in Figure 7.8.

Invariancy regions have a very intuitive interpretation for portfolio managers and financial analysts as inside each invariancy region the portfolio composition is fixed. By fixed composition we mean that the pool of assets included in the portfolio remains unchanged while asset weights can vary. For instance, on the invariancy region π_1 in Figure 7.7(b) the optimal partition is

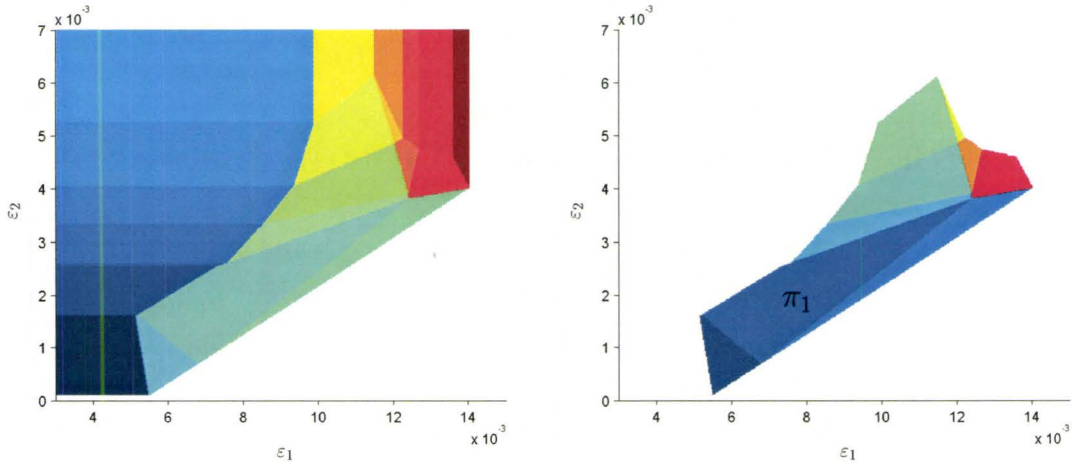


Figure 7.7: Invariancy Regions (a) and Invariancy Regions Corresponding to the Pareto Efficient Solutions (b) for the Mean-Variance Portfolio Optimization Problem with Transaction Cost.

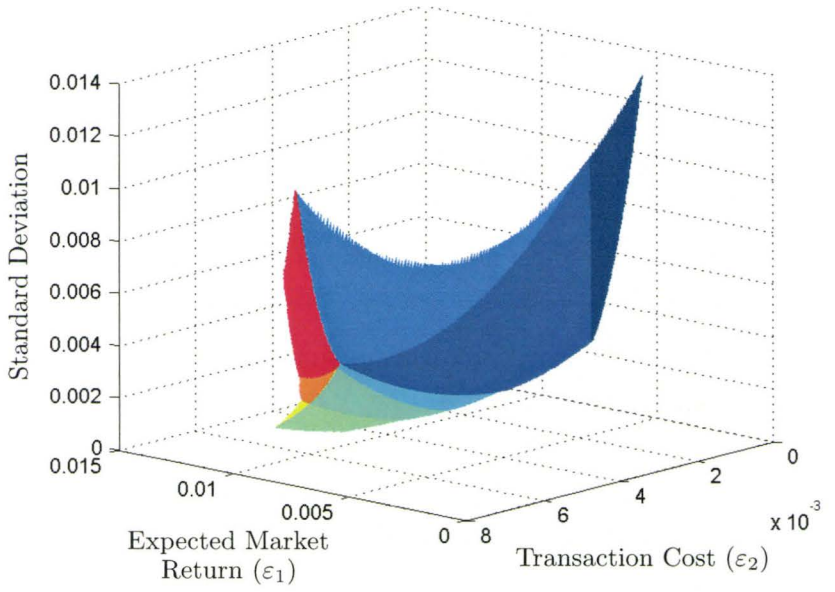


Figure 7.8: The Pareto Efficient Surface for the Mean-Variance Portfolio Optimization Problem with Transaction Cost.

$\mathcal{NNBBNBBN}$ which means that the portfolio is composed of securities 3, 4, 6 and 7. The functional form of the Pareto front on the invariancy region π_1 is $f_3 = 0.1 - 0.4f_1 - 23.7f_2 + 13.4f_1^2 + 11999.4f_2^2 - 621.9f_1f_2$.

7.4 Robust Mean-Variance Optimization

One of the common criticisms of mean-variance optimization is its sensitivity to return estimates. As the consequence of that fact, small changes in the return estimates can result in big shifts of the portfolio weights x. One of the solutions to this problem is *robust optimization*, which incorporates uncertainties into the optimization problem. For a review of the robust optimization applied to portfolio management we refer the reader to [47].

Mathematical formulation of the robust mean-variance optimization problem presented in [22] and described in Section 6.1.1.3 is the following:

$$\begin{array}{ll} \min & -\hat{r}^T x + \kappa \|\Theta^{1/2} x\| + \lambda x^T Q x \\ \text{s.t.} & \sum_{i=1}^{n} x_i = 1 \\ & x \geq 0, \end{array} \tag{7.4.1}$$

where $\hat{r}$ is the vector of expected returns, Θ is the covariance matrix of estimated expected returns, Q is the covariance matrix of returns, κ is the estimation error aversion, and λ is the risk aversion.

Formulation (7.4.1) is a parametric SOCO problem with two parameters κ and λ. Preliminary results on parametric SOCO are discussed in Chapter 8. If we look at it in the multiobjective sense, it is the problem of maximizing expected return, minimizing risk (variance of returns) and minimizing estimation error for the expected return. The problem formulation emphasizes the differences between the true, the estimated, and the actual Markowitz efficient frontiers [22].

To demonstrate the influence that sensitivities in the return estimates can potentially have on the portfolio selection, Ceria [21] presented a simple portfolio consisting of three assets. Table 7.6 shows expected returns for the two estimates and standard deviations for the assets. As Table 7.6 also shows, completely different portfolio weights can be obtained while optimizing the portfolio with

Table 7.6: Expected Returns and Standard Deviations with Correlations = 20% for Robust Mean-Variance Optimization, Optimal Weights for Two Portfolios.

Security	r^1	r^2	σ
Asset 1	7.15%	7.16%	20%
Asset 2	7.16%	7.15%	24%
Asset 3	7.00%	7.00%	28%

Security	Portfolio A	Portfolio B
Asset 1	38.1%	84.3%
Asset 2	69.1%	15.7%
Asset 3	0.0%	0.0%

expected return estimates r^1 and r^2. Taking r^1 as the estimate of the expected returns, we solve the multiobjective problem (7.4.1) to find all possible trade-offs between the three objectives – maximizing expected return, minimizing variance and minimizing estimation error.

As $x^TQx \leq \sigma_1^2$ ($Q = RR^T$) and $\|\Theta^{1/2}x\| = \sqrt{x^T\Theta x} \leq \sigma_2$, we can rewrite problem (7.4.1) in the form:

$$\begin{array}{ll} \min & -\hat{r}^Tx + \lambda_1 u_0 + \lambda_2 v_0 \\ \text{s.t.} & \sum_{i=1}^n x_i = 1 \\ & x \geq 0 \\ & \Theta^{1/2}x - u = 0 \\ & R^Tx - v = 0 \\ & (u_0, u) \in \mathcal{K}_q, \ (v_0, v) \in \mathcal{K}_q, \end{array} \tag{7.4.2}$$

where parameters $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ and $\mathcal{K}_q$ is the second-order cone. Parametric problem (7.4.2) represents the weighting method for multiobjective optimization.

Formulating the parametric problem corresponding to the ε-constrained method for multiobjective optimization we get:

$$\begin{array}{ll} \min & -\hat{r}^Tx \\ \text{s.t.} & \sum_{i=1}^n x_i = 1 \\ & x \geq 0 \\ & \Theta^{1/2}x - u = 0 \\ & R^Tx - v = 0 \\ & u_0 = \varepsilon_1 \\ & v_0 = \varepsilon_2 \\ & (u_0, u) \in \mathcal{K}_q, \ (v_0, v) \in \mathcal{K}_q, \end{array} \tag{7.4.3}$$

where parameters $\varepsilon_1 \geq 0$ and $\varepsilon_2 \geq 0$, and Θ is the identity matrix.

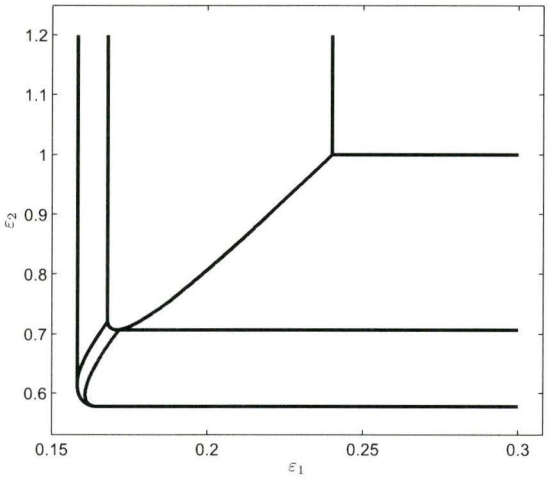


Figure 7.9: The Invariancy Regions for the Robust Mean-Variance Portfolio Optimization Problem.

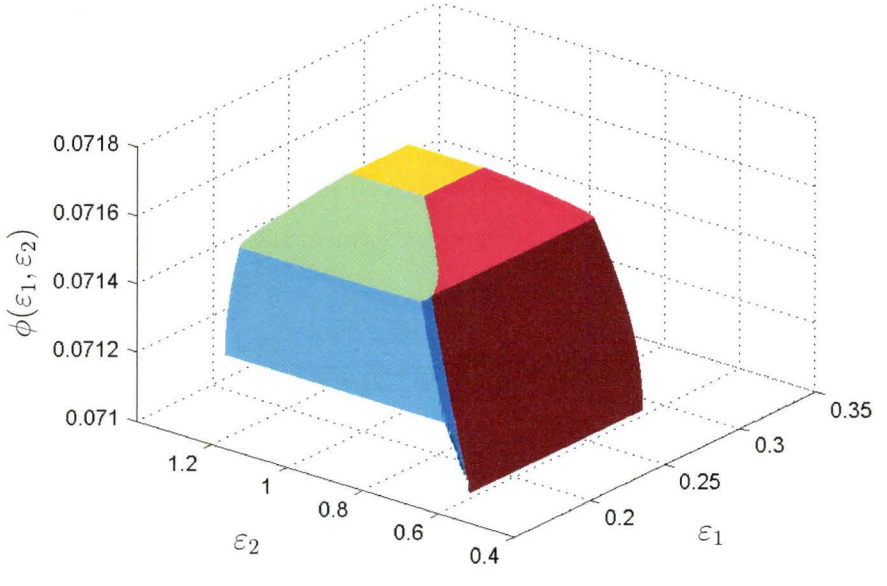


Figure 7.10: The Optimal Value Function for the Robust Mean-Variance Portfolio Optimization Problem.

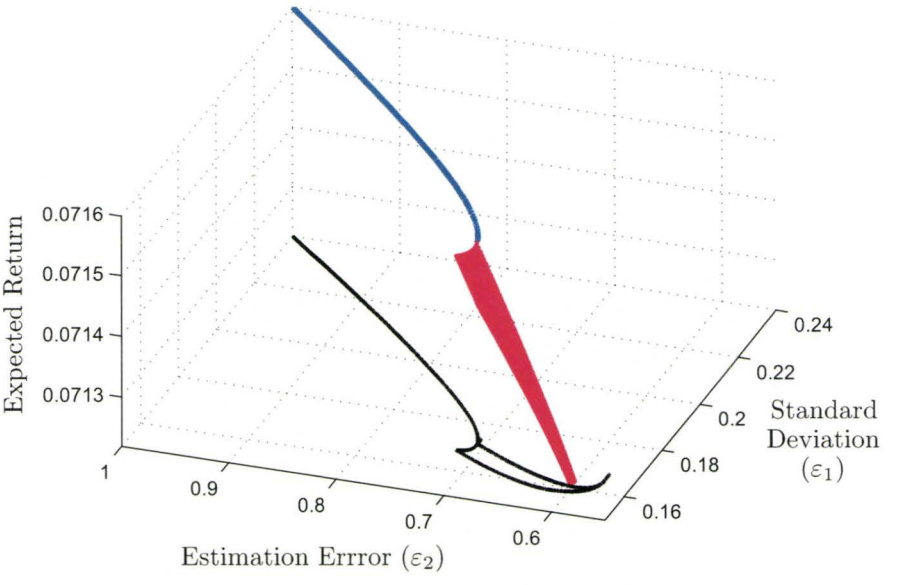


Figure 7.11: The Pareto Efficient Surface for the Robust Mean-Variance Portfolio Optimization Problem.

The optimal value function of the parametric SOCO formulation (7.4.3) with parameters $(\varepsilon_1, \varepsilon_2)$ in the constraints is shown in Figure 7.10. The corresponding invariancy regions are presented by Figure 7.9. To identify the invariancy regions that correspond to Pareto efficient solutions we need to restrict our attention to the regions where the second order conic blocks u and v belong to the subsets $\mathcal{R}$ or $\mathcal{T}$ of the optimal partition. Those invariancy regions and the corresponding Pareto efficient surface is shown in Figure 7.11.

7.5 Sparse and Stable Markowitz Portfolio Frontiers

In [16] a sparse mean-variance portfolio construction problem is considered. The authors consider N securities and denote their returns at time t by r_{it} with $i = 1, \ldots, N$ indexing securities and $t = 1, \ldots, T$ indexing historical time periods. Vector $r_t = (r_{1t}, \ldots, r_{Nt})^T$ is the vector of returns at time t. Assuming

stationarity of returns, the vector of expected returns is $\mathbb{E}[r_t] = r$ and the covariance matrix Q is computed as $Q = \mathbb{E}[(r_t - r)(r_t - r)^T] = \mathbb{E}[r_t r_t^T] - rr^T$. A portfolio is defined as a vector of weights $x = (x_1, \ldots, x_N)^T$ in assets. For the target value of portfolio return being equal to r_P, the mean-variance optimization problem is defined in [16] as:

$$\begin{array}{ll} \min\limits_{x} & \frac{1}{2} x^T Q x \\ \text{s.t.} & r^T x = r_P \\ & \sum_{i=1}^{N} x_i = 1, \end{array} \tag{7.5.1}$$

where $r = (r_1, \ldots, r_N)^T$ is computed as the sample average $r_i = \frac{1}{T} \sum_{t=1}^{T} r_{it}$.

The regularized variant of problem (7.5.1) with ℓ_1-norm penalty is defined as follows:

$$\begin{array}{ll} \min\limits_{x} & \frac{1}{2} x^T Q x + \lambda \|x\|_1 \\ \text{s.t.} & r^T x = r_P \\ & \sum_{i=1}^{N} x_i = 1, \end{array} \tag{7.5.2}$$

and solved in [16] by a generalized homotopy/LARS (Least Angle Regression) algorithm for a fixed value of r_P. For detailed description of regularized optimization we refer the reader to Section 6.2.

Optimization problem (7.5.2) is a bi-parametric QO problem. Let us add constraints $Ax = b$, $x \geq 0$ to the formulation and write it in the following form:

$$\begin{array}{ll} \min\limits_{x} & \frac{1}{2} x^T Q x + \lambda \sum_{i=1}^{N} |x_i| \\ \text{s.t.} & r^T x = \varepsilon \\ & \sum_{i=1}^{N} x_i = 1 \\ & Ax = b \\ & x \geq 0, \end{array} \tag{7.5.3}$$

where (λ, ε) are the problem parameters. We allow for general linear equality and inequality constraints in the form $Ax = b$, $x \geq 0$ as compared to the single linear equality constraint $\sum_{i=1}^{N} x_i = 1$ considered in [16]. Problem (7.5.3) is a multiobjective optimization problem with three objectives corresponding to maximizing portfolio sparsity, maximizing expected return and minimizing portfolio variance.

We make the following alterations:

1) Formulation (7.5.3) allows more general constraints than (7.5.2);
2) Formulation (7.5.3) treats expected return as the third problem objective in addition to minimizing portfolio variance and maximizing its sparsity, thus allows the variation of parameter ε as opposed to the constant value of r_p in formulation (7.5.2);
3) Parametric optimization problem (7.5.3) can be solved with Algorithm 2.2 that is described in Section 2.3.

We use data for the vector of expected returns r and the covariance matrix Q from Table 7.1. For the ease of exposition we solve the ε-constrained formulation of the multiobjective problem (7.5.3):

$$\begin{array}{ll} \min\limits_{x} & \frac{1}{2}x^T Q x \\ \text{s.t.} & r^T x = \varepsilon_1 \\ & \sum_{i=1}^{N} |x_i| = \varepsilon_2 \\ & \sum_{i=1}^{N} x_i = 1 \\ & x \geq -0.2, \end{array} \tag{7.5.4}$$

where the constraint $x \geq -0.2$ allows for 20% short sales. The constraint $\sum_{i=1}^{N} |x_i| = \varepsilon_2$ can be written in the norm notation as $\|x\|_1 = \varepsilon_2$. In the actual computation problem (7.5.4) was re-written in the standard form by splitting the variables $x = \overline{x} - \underline{x}$, with $\overline{x}, \underline{x} \geq 0$ in order to replace the absolute value function $|\cdot|$ by linear terms.

The invariancy regions for the parametric QO problem (7.5.4) are shown in Figure 7.12. We display the portfolio compositions next to the invariancy regions in Figure 7.12. The plus sign denotes long positions of the corresponding assets, while the minus sign denotes short positions. For instance, portfolio composition +4-5 denotes that the portfolio is composed of assets 4 and 5 with long position in asset 4 and short position in asset 5. When the portfolio composition is changing on the transition lines or transition points between invariancy regions, we displayed the corresponding composition next to the line. All possible portfolio compositions identified by regularized optimization are shown in Figure 7.12.

The Pareto front which is a subset of the optimal value function for the bi-parametric QO problem (7.5.4) is plotted in Figure 7.13. As we are interested in the cardinality of the portfolio instead of the objective function value $\|x\|_1$, the Pareto front was also plotted in the cardinality–return–standard deviation space in Figure 7.14. Finally, we plot the same objective value function from Figure 7.14 in the mean–standard deviation space in Figure 7.15. The family of efficient frontiers in Figure 7.15 are the frontiers corresponding to different cardinalities of the vector of optimal weights x^*. The cardinality of x^* in its turn is equal to the cardinality of set $\mathcal{B}$ of the optimal partition on that invariancy region.

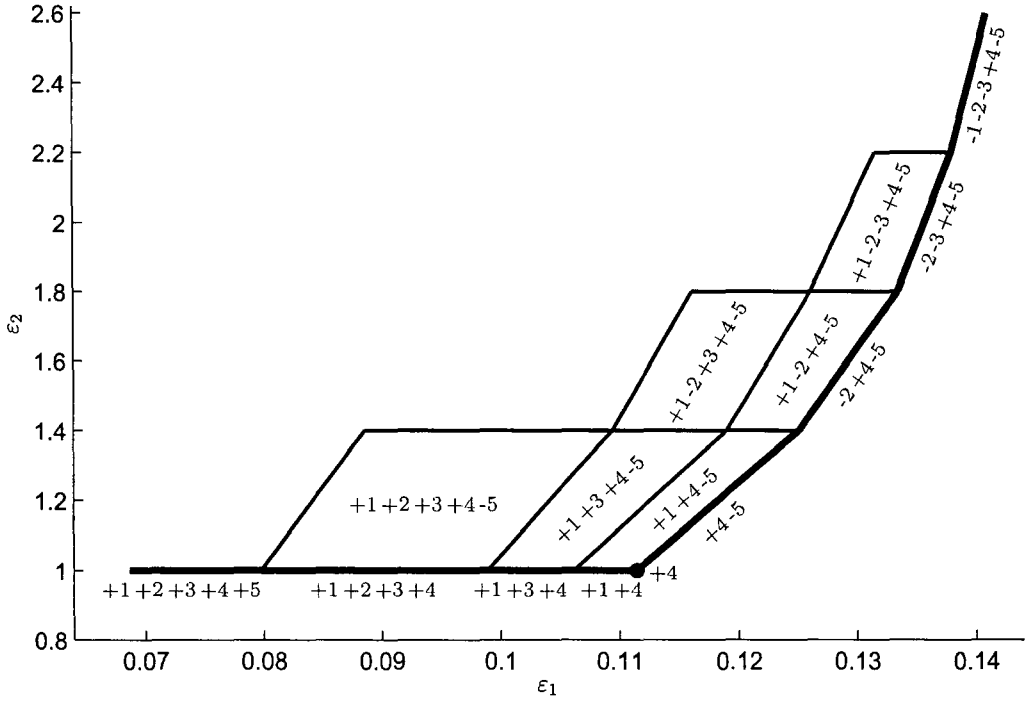


Figure 7.12: The Invariancy Regions for the Mean-Variance Problem with Market Risk and Sparsity Constraint.

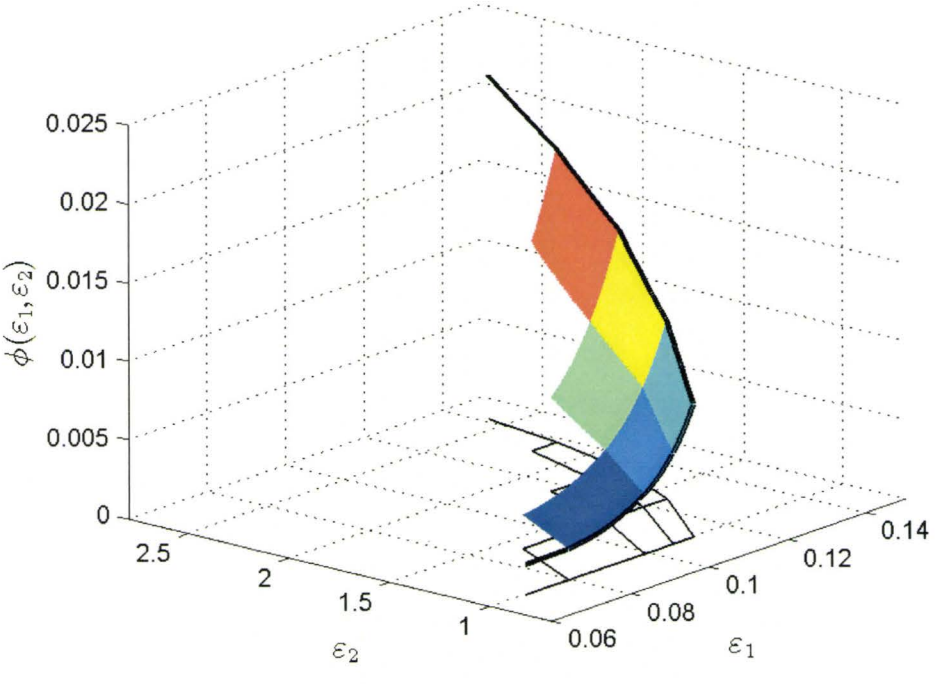


Figure 7.13: The Pareto Front and Optimal Value Function for the Mean-Variance Problem with Market Risk and Sparsity Constraint.

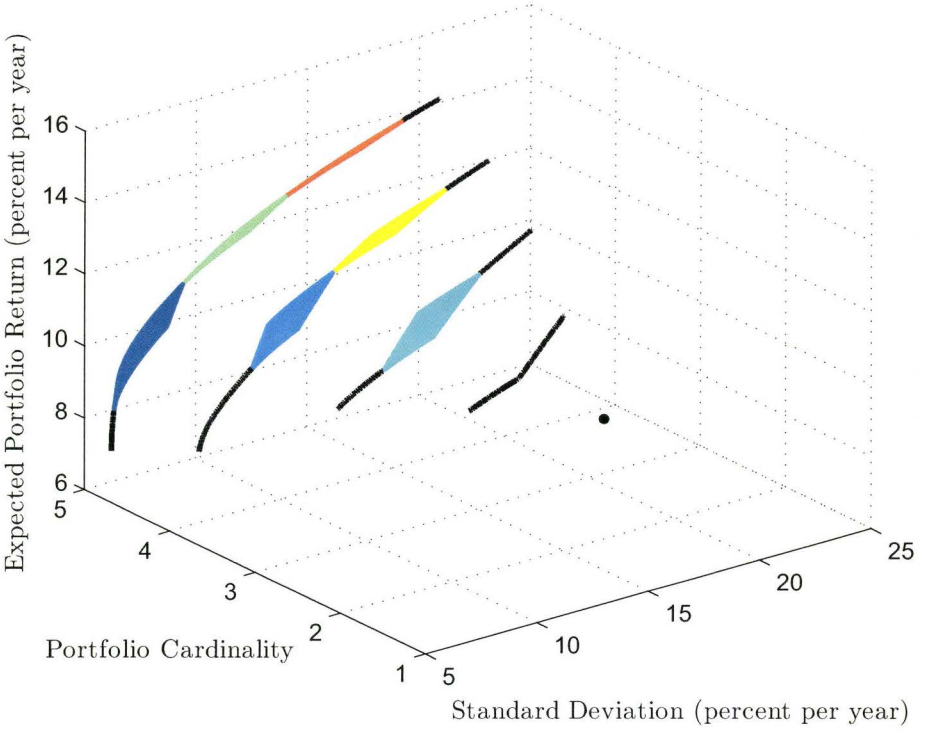


Figure 7.14: The Efficient Portfolio Frontier for the Mean-Variance Problem with Market Risk and Sparsity Constraint.

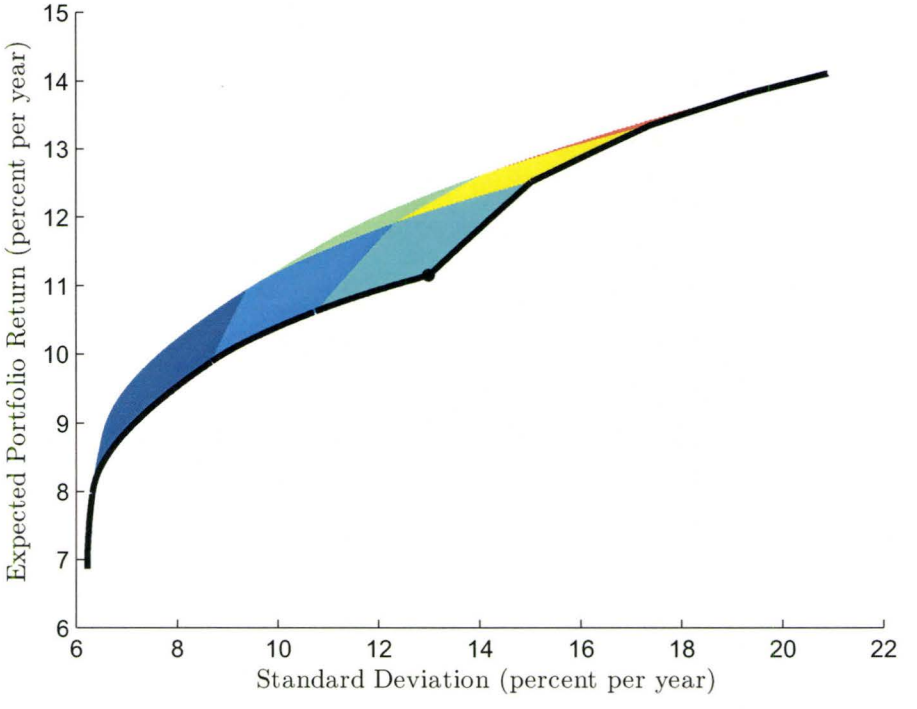


Figure 7.15: The Markowitz Efficient Portfolio Frontier for the Mean-Variance Problem with Market Risk and Sparsity Constraint.

7.6 Parametric Portfolio Replication

We use the portfolio replication framework, developed in Section 6.2, and consider a block of 15,000 variable annuity policies with a mixture of minimum guaranteed death benefits sold over a ten year period. The total annual cash flows comprising of both benefits paid and fees collected are projected over a 20-year time horizon. The economic scenario set used for the projection consists of 500 stochastic scenarios which are a mixture of risk neutral and real world scenarios.

Based on the characteristics of the liability, a tradable universe of 369 instruments is identified consisting of zero coupon bonds, index forwards and European index options on the indices, and swaptions. The maturities and strikes for the instruments are chosen to span the 20-year time period. Annual settlement cash flows for the instruments in the tradable universe are projected over a 20-year time horizon for each of the 500 economic scenarios used in the liability cash flow projection.

Utilizing the notation from Section 6.2, we obtain $S = 500$, $T = 20$, $N = 369$, $\omega_l^t = p_l = 1/S$ and matrix C with vector c_0 are given. We have used MATLAB for implementing portfolio replication formulations. The CPLEX [74] and the MOSEK [105] optimization solvers were called from MATLAB to find numerical solutions of the optimization problems.

Let us find the replicating portfolio that best matches the values of the liability cash flows at every time step. In this case, the 10,000 (20 years $\times$ 500 scenarios per year) liability cash flows cannot be replicated exactly by a portfolio of 369 instruments. We use the quadratic discrepancy measure $f(x) = \|Ax - b\|_2$. In addition, a linear constraint $v^T x = v_P$ on the total portfolio value is present. Now, based on formulation (6.2.14), we can present the optimization problem that we need to solve:

$$\begin{array}{ll} \min\limits_{x} & \|Ax - b\|_2^2 \\ \text{s.t.} & \|Dx\|_1 \leq \varepsilon \\ & v^T x = v_P. \end{array} \tag{7.6.1}$$

Similarly to [50], we reformulate problem (7.6.1) as a parametric quadratic optimization problem by splitting variable x into positive and negative parts:

$$x = \overline{x} - \underline{x},\ \overline{x} \geq 0,\ \underline{x} \geq 0.$$

Now, $\|Dx\|_1 = d^T\overline{x} + d^T\underline{x}$ and

$$\begin{array}{ll} \min\limits_{\overline{x},\underline{x}} & \|A(\overline{x} - \underline{x}) - b\|_2^2 \\ \text{s.t.} & d^T\overline{x} + d^T\underline{x} \leq \varepsilon, \\ & v^T\overline{x} - v^T\underline{x} = v_P, \\ & \overline{x} \geq 0, \\ & \underline{x} \geq 0. \end{array} \tag{7.6.2}$$

Rewriting problem (7.6.2) as a quadratic optimization problem in the standard form, we get:

$$\begin{array}{ll} \min\limits_{\tilde{x},t} & c^T\tilde{x} + \frac{1}{2}\tilde{x}^T B\tilde{x} + \frac{1}{2}b^T b \\ \text{s.t.} & \bar{d}^T\tilde{x} + t = \varepsilon, \\ & u^T\tilde{x} = v_P, \\ & \tilde{x} \geq 0,\ t \geq 0, \end{array} \tag{7.6.3}$$

where

$$\tilde{x} = \begin{bmatrix} \overline{x} \\ \underline{x} \end{bmatrix},\ c = \begin{bmatrix} -A^T b \\ A^T b \end{bmatrix},\ u = \begin{bmatrix} v \\ -v \end{bmatrix},\ \bar{d} = \begin{bmatrix} d \\ d \end{bmatrix},\ B = \begin{bmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{bmatrix}.$$

In the parametric quadratic optimization problem (7.6.3), t is a slack variable used to convert the inequality constraint into equality. The constant term $b^T b$ is ignored during optimization, but is added to the optimal value function.

For our computational experiments, we solve two parametric problems. The first one is problem (7.6.3) with $\bar{d} = (\mathbf{1}_N, \mathbf{1}_N)^T$. The second one is problem (7.6.3) with $\bar{d} = (\frac{1}{|x^*|}, \frac{1}{|x^*|})^T$, where x^* is the optimal solution of problem (7.6.1) without the regularization constraint $\|Dx\|_1 \leq \varepsilon$.

First, we restrict the total number of units traded, i.e., $d_i = 1$ for all i in equation (7.6.1) and its standard form (7.6.3), and minimize the expected squared deviation of the cash flows, i.e., minimize $f(x)$ as defined in equation (6.2.2). Solving parametric problem (7.6.3) with ε as the parameter produces an efficient frontier that shows the optimal trade-off between trade volume and in-sample

replication error, see Figures 7.16 and 7.17. Locations of optimal solutions on the efficient frontier are determined by parameter ε, which equals to the total units traded $\|x\|_1$ at the corresponding location. Figure 7.16 shows the Pareto efficient frontier, which is also the optimal value function $\phi(\varepsilon)$, for the area of interest. In Figure 7.16 the transition points are the dots and the invariancy intervals are the curves between transition points. Figure 7.17 shows the whole frontier. The horizontal dashed lines in Figures 7.16 and 7.17 correspond to solutions without trading penalty.

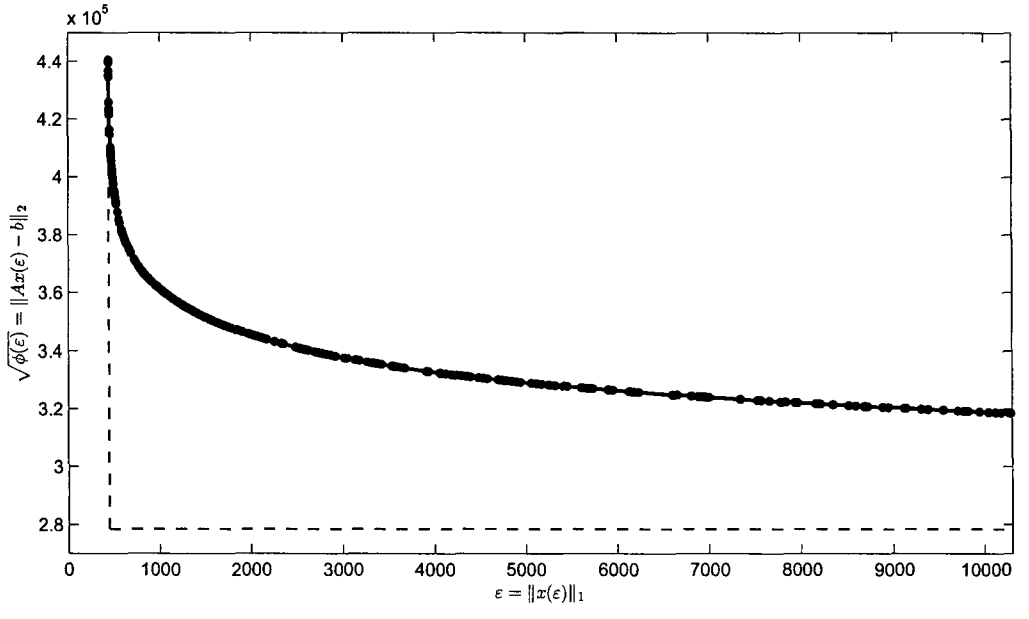


Figure 7.16: The Optimal Value Function and the Invariancy Intervals for the Portfolio Replication Problem.

Figures 7.18, and 7.19 show the optimal solution cardinality vs. trading budget, and the optimal solution cardinality vs. in-sample replication error, respectively. As the trading budget increases, the in-sample replication error decreases. Observe that increasing the budget causes more instruments to be included in the replicating portfolio, see Figure 7.18. Moreover, this increase is accompanied by improved in-sample performance, see Figure 7.19.

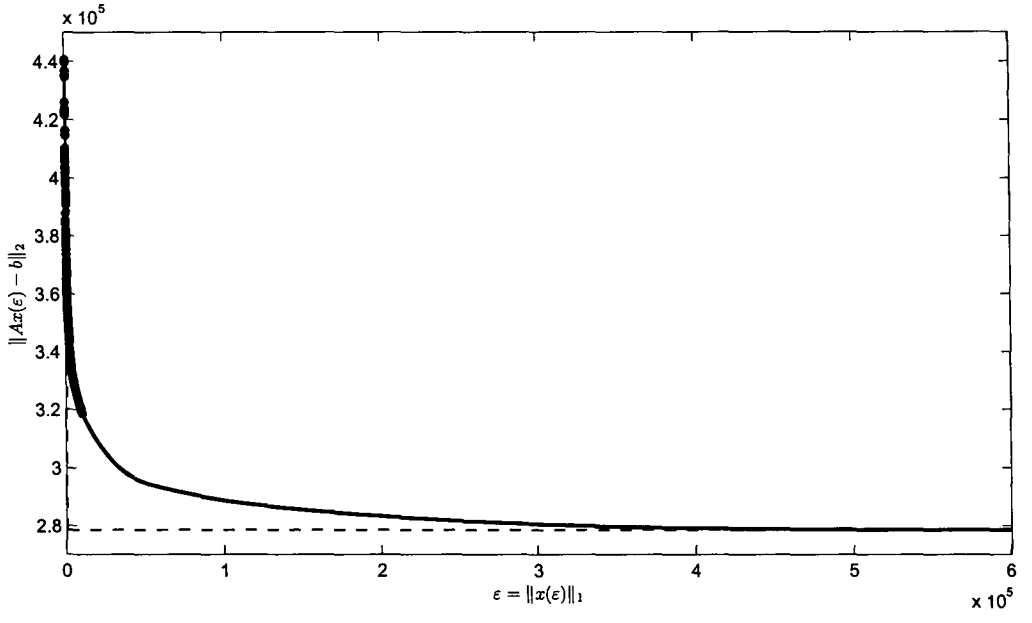


Figure 7.17: The Efficient Frontier Between Replication Error and Trading Penalty.

The optimal portfolio composition is shown in Figure 7.20, where the horizontal axis depicts instrument number and the vertical axis shows the replication mismatch $\|Ax - b\|_2$.

The second formulation, that we solved, restricts the total number of units traded to $d_i = \frac{1}{|x_i^*|}$ for all i, where x^* is the optimal solution of problem (7.6.1) without the regularization constraint $\|Dx\|_1 \leq \varepsilon$. For the obtained value of vector $\bar{d} = (d, d)^T = (\frac{1}{|x^*|}, \frac{1}{|x^*|})^T$, we optimized the parametric formulation (7.6.3). Solving parametric problem (7.6.3) with the *weighted* ℓ_1 *regularization* constraint parameterized by ε produces another Pareto efficient frontier. Table 7.7 shows an extract from solver output for the parametric problem, where only the first five invariancy intervals are shown. The output includes invariancy intervals and transition points corresponding to parameter ε; portfolio composition, namely subset $\mathcal{B}$ of the optimal partition with indexes of the selected instruments; and the scaled optimal value function $\phi(\varepsilon)$ that describes dependence of the replica-

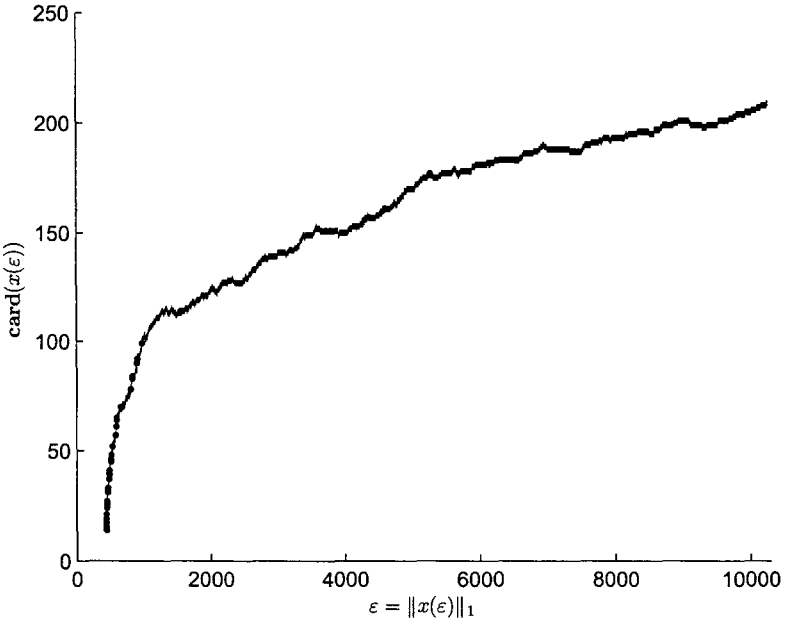


Figure 7.18: Optimal Solution Cardinality vs. $\|x\|_1$ for the Portfolio Replication Problem.

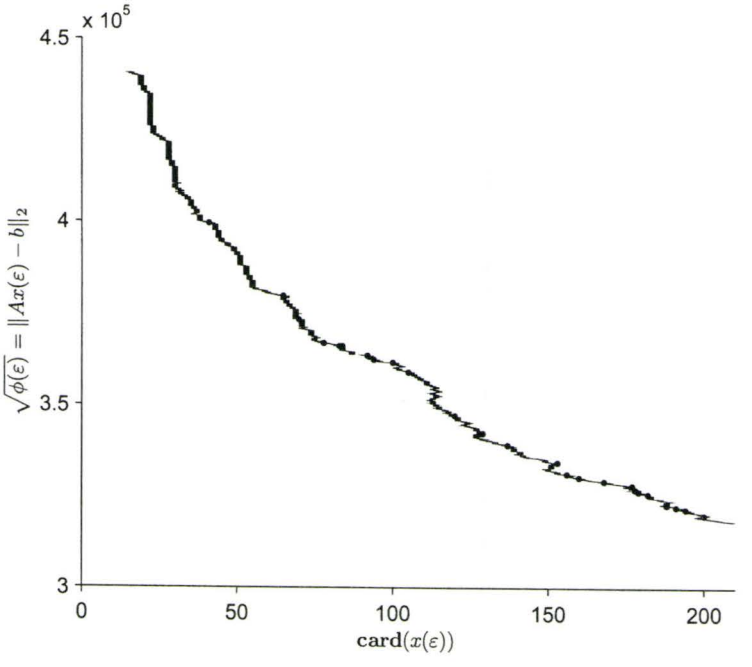


Figure 7.19: Optimal Solution Cardinality vs. $\|Ax - b\|_2$ for the Portfolio Replication Problem.

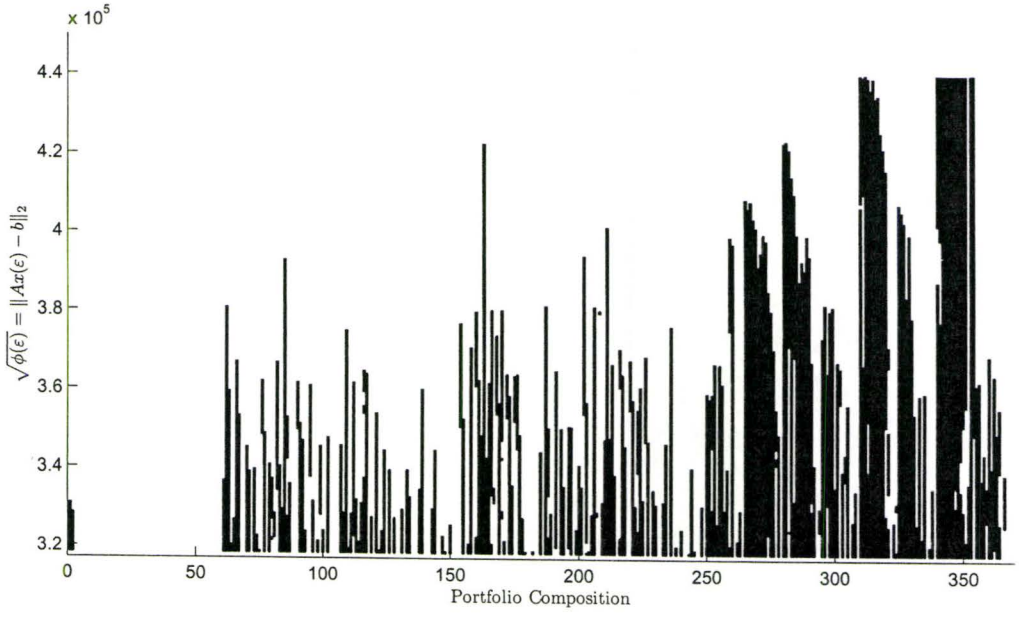


Figure 7.20: The Composition of the Optimal Replicating Portfolio.

Table 7.7: Parametric Solver Output for the Portfolio Replication Problem with Weighted ℓ_1 Regularization.

type	ε_ℓ	ε_u	$\mathcal{B}$	$\mathcal{T}$	$\phi(\varepsilon)$
tr. point	+0.05650	+0.05650	15	289	23208.70490
inv. interv	+0.05650	+0.40684	15 289		$18897.49\varepsilon^2$-31075.68ε+24904.05
tr. point	+0.40684	+0.40684	8 15 289	8	15389.08765
inv. interv	+0.40684	+0.82583	8 15 289		$8563.38\varepsilon^2$-22679.70ε+23198.71
tr. point	+0.82583	+0.82583	8 15 289	3	10309.30802
inv. interv	+0.82583	+0.83121	3 8 15 289		$6796.10\varepsilon^2$-19760.74ε+21993.43
tr. point	+0.83121	+0.83121	3 8 15 289	287	10263.549615
inv. interv	+0.83121	+0.99040	3 8 15 287 289		$6451.42\varepsilon^2$-19187.75ε+21755.28
tr. point	+0.99040	+0.99040	3 8 15 287 289	284	9079.89343
inv. interv	+0.99040	+1.13110	3 8 15 284 287 289		$4457.99\varepsilon^2$-15240.04ε+19800.83
tr. point	+1.13110	+1.13110	...		

tion error on the trading penalty ε in functional form.

Now, we can compare results produced by parametric optimization problems with $d_i = 1$ and $d_i = \frac{1}{|x_i^*|}$ to the exact solution of the cardinality-constrained problem, when the regularization constraint $\|Dx\|_1 \leq \varepsilon$ in (7.6.1) is replaced by the cardinality constraint $\text{card}(x) = K$. The cardinality-constrained problem was formulated as

$$\begin{array}{ll} \min\limits_{x,y} & \|Ax - b\|_2^2 \\ \text{s.t.} & v^T x = v_p \\ & \sum_{i=1}^N y_i = K \\ & e_\ell \cdot y \leq x \leq e_u \cdot y \\ & y = \{0, 1\}, \end{array} \tag{7.6.4}$$

where y is the binary variable, e_ℓ and e_u are appropriately chosen lower and upper bounds on instrument positions.

The mixed-integer QO problem (7.6.4) was solved for $K = 15, 25, \ldots, 145$ with CPLEX and the running time for each value of K was restricted to 24 hours. The parametric solution with an appropriate cardinality K obtained by solving problem (7.6.3) was used as the initial seed.

Comparison of the solutions produced by parametric optimization with integer solutions is demonstrated by Figure 7.21. Figure 7.21 shows dependance of the replication error measured by squared deviations on the number of in-

struments in the replicating portfolio, i.e. portfolio cardinality. Three cases are shown: when the parametric problem is solved for $d_i = 1$; for $d_i = \frac{1}{|x_i^*|}$, where x_i^* is the number of units of instrument i held in the optimal replicating portfolio for the problem without trading penalty; and for the solution of the mixed-integer problem (7.6.4) with cardinality constraint. In Figure 7.21 those cases are referred to as ℓ_1-norm regularized heuristic, ℓ_1-norm regularized weighted heuristic and mixed-integer optimization, respectively.

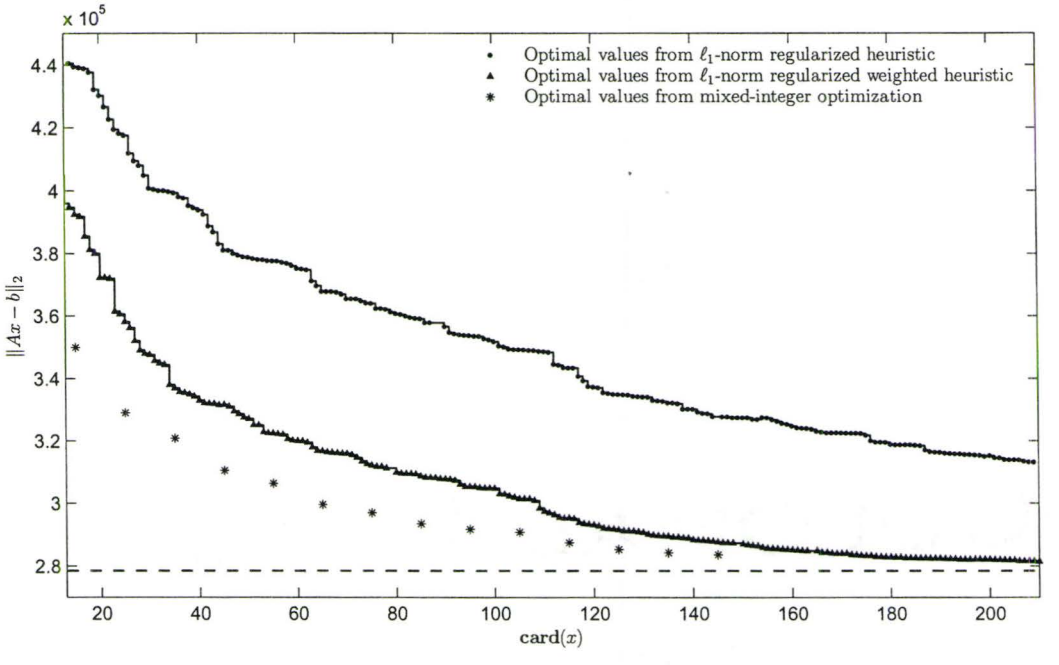


Figure 7.21: Replication Error vs. Solution Cardinality for Solutions Produced by Parametric Optimization and Mixed-Integer Optimization.

Now, we can compare the results obtained by mixed-integer optimization for the instrument selection, with the results obtained by the parametric optimization algorithm. It turns out that the quality of the parametric optimization solutions with $d_i = \frac{1}{|x_i^*|}$ is comparable to the results of mixed-integer optimization, if the quality measure is defined by the replication error. For the data set that we used for testing, parametric optimization can produce replication er-

ror vs. solution cardinality efficient frontier that approximates the true frontier computed by integer optimization quite well. However, the parametric optimization frontier is computed in minutes, while the integer optimization frontier is computed in days.

7.7 Expected Shortfall Optimization for Different Quantiles

The portfolio credit risk optimization formulations described in Section 6.3 treat quantile level α for quantile-based risk measures as fixed. Computational results are reported in Appendix A, where different formulations were optimized for a number of quantile levels, see Figure A.7. In many cases, risk measures should be optimized for a range of quantiles, and the question of interest is how expected shortfall $\mathrm{ES}_\alpha(\mathcal{L}(x))$ in the optimum depends on quantile level α.

Let us consider LO problem (6.3.24) of minimizing expected shortfall with LLN approximation. We define parameter λ as $\lambda = \frac{1}{1-\alpha}$ and rewrite problem (6.3.24) as parametric optimization problem:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^N,\ u\in\mathbb{R}^M,\ \ell\in\mathbb{R}} & \ell + \lambda\dfrac{1}{M}\displaystyle\sum_{l=1}^{M} u_l \\ \text{s.t.} & u_l \geq \mu_l(x) - \ell,\ u_l \geq 0,\ l = 1,\ldots,M \\ & x \in \Omega, \end{array} \tag{7.7.1}$$

where parameter λ appears in the objective.

Formulation (7.7.1) is a uni-parametric LO problem and it can be solved using the techniques described in Section 2.2. In practice, quantile levels that are considered by credit risk practitioners range from 99% to 99.97%, that defines the interval of the parameter $\lambda = [100, 3333.34]$, where parametric problem (7.7.1) need to be solved. A similar parametric problem can be formulated for the expected shortfall optimization problem (6.3.19) with MC-sampling.

The piecewise linear optimal value function $\phi(\lambda)$ for the parametric LO problem (7.7.1) is plotted in Figure 7.22. As the real interest is in getting the

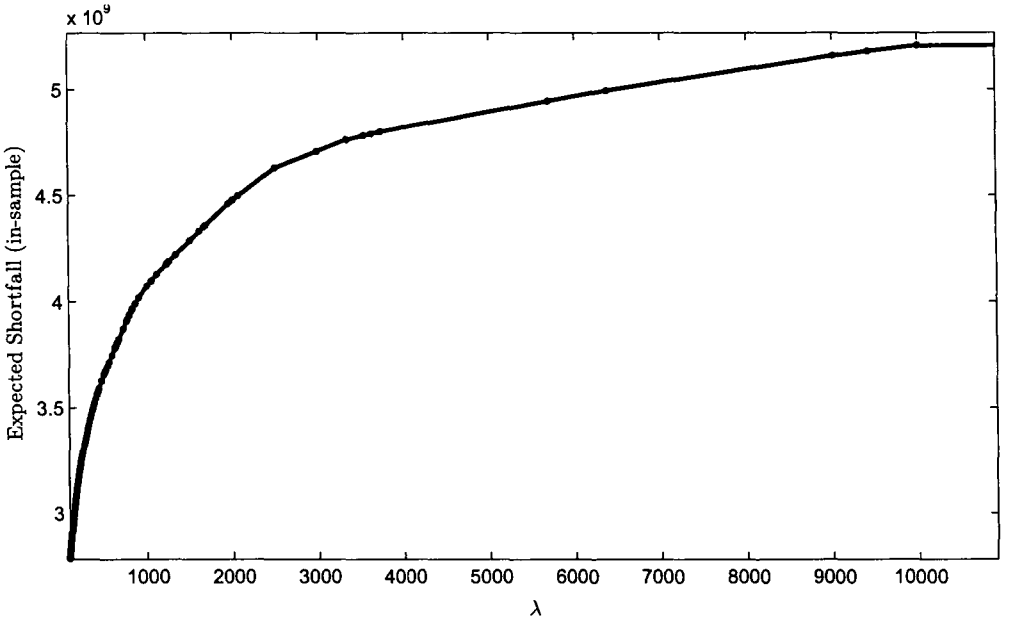


Figure 7.22: The Optimal Value Function for Expected Shortfall Optimization with Different Quantiles.

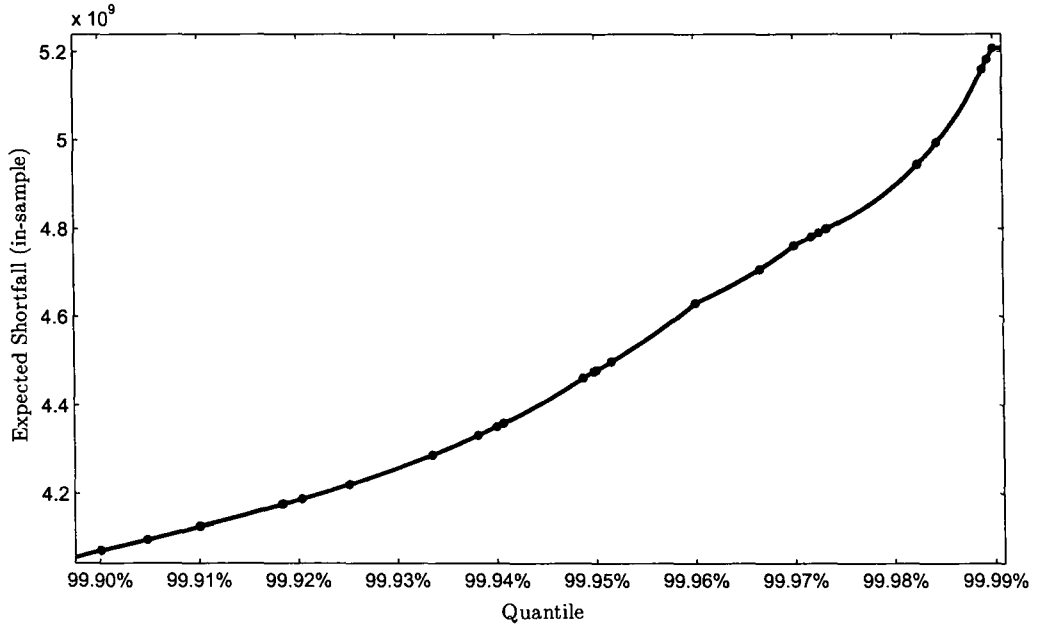


Figure 7.23: The Expected Shortfall vs. Quantile for Expected Shortfall Optimization with Different Quantiles.

dependence of the expected shortfall ES on the quantile level α, we have used the optimal value function $\phi(\lambda)$ to compute the piecewise function $\mathrm{ES}(\alpha)$. The expected shortfall as a function of quantile level is shown in Figure 7.23. For instance, on the interval $\alpha = (0.9952, 0.9960)$, the expected shortfall depends on α as

$$\mathrm{ES} = 3873898968 + \frac{302551}{1-\alpha}.$$

Chapter 8

Discussions on Parametric Second-Order Conic Optimization

Parametric SOCO is a natural extension of parametric analysis for LO and QO. As we point out in Section 3.3, parametric SOCO allows solving multiobjective quadratic optimization problems with more than one quadratic objective. The *optimal basis approach* to parametric optimization in LO cannot be directly generalized to parametric optimization in SOCO [152]. In contrast, it is promising to generalize the *optimal partition approach* of parametric LO and QO to SOCO. We describe ideas and preliminary results related to parametric SOCO in this chapter.

The standard form SOCO problem is defined in Section 1.1.4. Primal problem ($SOCP$) and dual problem ($SOCD$) are specified by equations (1.1.10) and (1.1.11), respectively. Before defining parametric SOCO formally, we describe the geometry of second-order (quadratic) cones. An extensive review of SOCO problems can be found in [4].

Unlike LO and QO, in SOCO we work with blocks of primal and dual variables, see the definition of ($SOCP$) and ($SOCD$) problems in Section 1.1.4. Those primal-dual blocks (x^i, s^i), $i = 1, \ldots, I$ of variables compose the decision vectors of SOCO problems $x = (x^1, \ldots, x^I)^T$ and $s = (s^1, \ldots, s^I)^T$, where $x^i, s^i \in$

$\mathbb{R}^{n_i}$. We also refer to cone $\mathcal{K}$ as a second-order cone when it is a product cone $\mathcal{K} = \mathcal{K}_q^1 \times \ldots \times \mathcal{K}_q^I$, where $x^i \in \mathcal{K}_q^i$, $i = 1, \ldots, I$. As a linear cone $\mathcal{K}_\ell^i$ is a one-dimensional quadratic cone $\mathcal{K}_q^i$ ($x_1^i \geq 0$), we treat linear variables as one-dimensional blocks. As before, $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$.

The bi-parametric SOCO problem in primal and dual form is expressed as:

$$(SOCP_{\lambda,\epsilon}) \qquad \begin{array}{rcll} \phi(\lambda,\epsilon) & = & \min & (c+\lambda\triangle c)^T x \\ & & \text{s.t.} & Ax = b + \epsilon\triangle b \\ & & & x \in \mathcal{K}, \end{array} \tag{8.0.1}$$

and

$$(SOCD_{\lambda,\epsilon}) \qquad \begin{array}{rll} \max & (b+\epsilon\triangle b)^T y \\ \text{s.t.} & A^T y + s = c + \lambda\triangle c \\ & s \in \mathcal{K}^*, \end{array} \tag{8.0.2}$$

where $A \in \mathbb{R}^{m\times n}$, $\text{rank}(A) = m$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ are fixed data; $x, s \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are unknown vectors; $\lambda, \epsilon \in \mathbb{R}$ are the perturbation parameters. Note that constraints $x \in \mathcal{K}$ and $s \in \mathcal{K}^*$ are replaced by $x_1^i \geq \|x_{2:n_i}^i\|_2$ and $s_1^i \geq \|s_{2:n_i}^i\|_2$, $i = 1, \ldots, I$ for computational purposes.

8.1 The Optimal Partition in SOCO

Algebraic representation of the optimal partition for SOCO problems is required for computational purposes. It will allow identification of invariancy intervals for parametric SOCO problems.

Yildirim [153] has introduced an optimal partition concept for conic optimization. He took a geometric approach in defining the optimal partition while using an algebraic approach is necessary for algorithm design. Although, the geometric approach has the advantage of being independent from the representation of the underlying optimization problem, it has some deficiencies. The major difficulty is extracting the optimal partition from a high-dimensional geometric object and, consequently, it is inconvenient for numerical calculations. In contrast, the algebraic approach is directly applicable for numerical implementation.

More recent study [13] provided the definition of the optimal partition that can be adapted to algebraic approach. We describe the details and compare the definitions of the optimal partition for SOCO in [153] (its algebraic form) and [13] in this section. Before defining the optimal partition for SOCO formally, we introduce the necessary concepts and notation. The interior and boundary of second-order cones are defined as follows.

Definition 8.1.1 (Interior of second-order cones) *The interior of second-order cone $\mathcal{K}_q \in \mathbb{R}^n$ is*

$$\operatorname{int} \mathcal{K}_q = \{x \in \mathcal{K}_q : x_1 > \|x_{2:n}\|_2\}.$$

Definition 8.1.2 (Second-order cone boundary) *The boundary of second-order cone $\mathcal{K}_q \in \mathbb{R}^n$ without the origin 0 is*

$$\operatorname{bd} \mathcal{K}_q = \{x \in \mathcal{K}_q : x_1 = \|x_{2:n}\|_2,\ x \neq 0\}.$$

Assuming strong duality, the *optimality conditions* for SOCO problems are:

$$\begin{aligned} Ax - b &= 0, \quad x \in \mathcal{K}, \\ A^T y + s - c &= 0, \quad s \in \mathcal{K}, \\ x \circ s &= 0, \end{aligned}$$

where the multiplication operation "$\circ$" is defined as $x \circ s = (x^1 \circ s^1, \ldots, x^I \circ s^I)^T$ and $x^i \circ s^i = ((x^i)^T s^i,\ x_1^i s_2^i + s_1^i x_2^i,\ \ldots,\ x_1^i s_{n_i}^i + s_1^i x_{n_i}^i)^T$.

Strict complementarity for SOCO problems [4] is defined as $x^i \circ s^i = 0$ and $x^i + s^i \in \operatorname{int} \mathcal{K}_q^i$, $i = 1, \ldots, I$. Interior point methods for SOCO produce maximally complementary solutions that maximize the number of strictly complementary blocks i.

With respect to its cone $\mathcal{K}_q^i$ each block x^i can be in one of three states:

- block x^i is in the interior of $\mathcal{K}_q^i$:

$$\operatorname{int} \mathcal{K}_q^i = \{\, x^i \in \mathcal{K}_q^i : x_1^i > \|x_{2:n_i}^i\|_2 \,\},$$

- block x^i is on the boundary of $\mathcal{K}_q^i$:

$$\text{bd } \mathcal{K}_q^i = \{ x^i \in \mathcal{K}_q^i : x_1^i = \|x_{2:n_i}^i\|_2 \text{ and } x^i \neq 0 \},$$

- block x^i equals 0:

$$x^i = 0.$$

The same results are valid for the dual blocks of variables $s^i \in (\mathcal{K}_q^i)^*$. As second-order cones are self-dual $\mathcal{K} = \mathcal{K}^*$, we are going to denote both primal and dual cones by $\mathcal{K}$ in the remainder of this chapter.

The optimal partition for SOCO has four sets, so it is a 4-partition $\pi = (\mathcal{B}, \mathcal{N}, \mathcal{R}, \mathcal{T})$ of the index set $\{1, 2, \ldots, I\}$. The four subsets are defined in [13] as:

$$\begin{aligned}
\mathcal{B} &= \{ i : x_1^i > \|x_{2:n_i}^i\|_2 \ (x^i \in \text{int } \mathcal{K}_q^i) \text{ for a primal optimal solution } x \}, \\
\mathcal{N} &= \{ i : s_1^i > \|s_{2:n_i}^i\|_2 \ (s^i \in \text{int } \mathcal{K}_q^i) \text{ for a dual optimal solution } (y, s) \}, \\
\mathcal{R} &= \{ i : x^i \neq 0 \neq s^i \ (x^i \in \text{bd } \mathcal{K}_q^i \text{ and } s^i \in \text{bd } \mathcal{K}_q^i) \\
&\qquad \text{for a primal-dual optimal solution } (x, y, s) \}, \\
\mathcal{T} &= \{ i : x^i = s^i = 0, \text{ or } s^i = 0 \text{ and } s^i \in \text{bd } \mathcal{K}_q^i, \text{ or } s^i = 0 \text{ and } x^i \in \text{bd } \mathcal{K}_i^q \\
&\qquad \text{for a primal-dual optimal solution } (x, y, s) \}.
\end{aligned}$$

Now we can state all possible configurations for primal-dual blocks of variables at optimality, those are summarized in Table 8.1 and serve as a basis for defining the optimal partition. Cases that are not geometrically possible, as those do not satisfy the optimality conditions, are shown as "×" in Table 8.1.

For the set $\mathcal{R}$ of the optimal partition it holds that $x^i \neq 0 \neq s^i$, and those blocks x^i and s^i lie on the boundary of $\mathcal{K}$ (i.e., $x_1^i = \|x_{2:n_i}^i\|_2 \neq 0$ and analogous relation holds for the dual). Let $(\bar{x}, \bar{y}, \bar{s})$ be a maximally complementary solution of problems $(SOCP)$ and $(SOCD)$ defined by (1.1.10)-(1.1.11), then as $x^i \circ s^i = 0$ we have

$$x^i \in \{ \alpha \bar{x}^i : \alpha \geq 0 \},$$

$$s^i \in \{ \beta(\bar{x}_1^i, -\bar{x}_{2:n_i}^i) : \beta \geq 0 \},$$

Table 8.1: Optimal Partition for SOCO.

s^i \ x^i	0	bd $\mathcal{K}_q^i$	int $\mathcal{K}_q^i$
0	$i \in \mathcal{T}$	$i \in \mathcal{T}$	$i \in \mathcal{B}$
bd $\mathcal{K}_q^i$	$i \in \mathcal{T}$	$i \in \mathcal{R}$	$\times$
int $\mathcal{K}_q^i$	$i \in \mathcal{N}$	$\times$	$\times$

is equivalent to the primal and dual blocks belonging to orthogonal boundary rays of the cone $\mathcal{K}_q^i$.

We can replace the definition of the set $\mathcal{R}$ of the optimal partition by:

$$\mathcal{R}(\bar{x}) = \{\, (i, \bar{x}^i) : x^i \neq 0 \neq s^i,\ x^i \in \{\alpha \bar{x}^i : \alpha \geq 0\},\ s^i \in \{\beta(\bar{x}_1^i, -\bar{x}_{2:n_i}^i) : \beta \geq 0\} \text{ for a primal-dual optimal solution } (x, y, s) \,\}.$$

Based on the results of Yildirim [153], we can alternatively define the optimal partition in algebraic form as $\pi_r = (\mathcal{B}, \mathcal{N}, \mathcal{R}(\bar{x}), \mathcal{T})$. The difference from the definition of π is that for primal-dual boundary blocks, it holds that $x^i \in$ a specific boundary ray of $\mathcal{K}_q^i$ and $s^i \in$ the orthogonal boundary ray of $\mathcal{K}_q^i$, instead of $x^i \in \text{bd}\,\mathcal{K}_q^i$ and $s^i \in \text{bd}\,\mathcal{K}_q^i$.

Comparing the two definitions of the optimal partition, π and π_r, it is worth to mention a couple of differences. When the optimal partition is defined as π, it partitions the index set $\{1, 2, \ldots, I\}$ of the blocks of variables. Consequently, it directly extends the definition of the optimal partition for QO (see Section 2.1) by adding the additional set $\mathcal{R}$ that corresponds to primal-dual optimal solutions being on the boundary of the cone, i.e., the case that does not exist for QO. In contrast, when the optimal partition is defined as π_r, it partitions not only the index set $\{1, 2, \ldots, I\}$, but also the space, as the set $\mathcal{R}(\bar{x})$ includes both indices of the blocks i and vectors $\bar{x}^i$ that define specific boundary rays. Definition of the optimal partition π_r is similar to the definition of the optimal partition for semidefinite optimization (SDO) in [60], which partitions the space and not the index set. Note that the real meaning of the partition set $\mathcal{R}(\bar{x})$ is that primal

and dual vectors should be on the boundary of the cone and belong to a specific ray on that boundary. If the optimal solution stays on the boundary, but moves to another boundary ray when the problem $(SOCP)$ is perturbed, the optimal partition π_r changes, while π remains invariant.

8.2 Bi-Parametric SOCO

Lets us consider the bi-parametric SOCO problem (8.0.1)-(8.0.2). We assume that the unperturbed problem $(SOCP_{0,0})$, where $\lambda = \epsilon = 0$, has non-empty primal and dual optimal solution sets and strong duality holds for it, i.e., the duality gap is zero. For now, we use the definition π_r of the optimal partition.

Similar to parametric QO in Chapter 2, we can transform the bi-parametric SOCO problem into a series of uni-parametric problems. For simplicity, let us assign $\epsilon = \lambda$. Moreover, let (x^*, y^*, s^*) be a maximally complementary optimal solution for $\lambda = 0$ with the optimal partition $\pi_r = (\mathcal{B}, \mathcal{N}, \mathcal{R}(x^*), \mathcal{T})$, the endpoints of the invariancy interval containing λ can be computed as:

$$\begin{aligned}\lambda_\ell \;=\; \min_{\lambda,x,y,s,\alpha,\beta} \; & \{\lambda : Ax - \lambda\triangle b = b,\; x_{\mathcal{B}\cup\mathcal{T}} \in \mathcal{K}_{\mathcal{B}\cup\mathcal{T}},\; x_{\mathcal{N}} = 0,\; x_{\mathcal{R}} = \alpha x^*_{\mathcal{R}},\; \alpha \geq 0,\\ & A^T y + s - \lambda\triangle c = c,\; s_{\mathcal{N}\cup\mathcal{T}} \in \mathcal{K}_{\mathcal{N}\cup\mathcal{T}},\; s_{\mathcal{B}} = 0,\; s_{\mathcal{R}} = \beta s^*_{\mathcal{R}},\; \beta \geq 0\},\end{aligned}$$

$$\begin{aligned}\lambda_u \;=\; \max_{\lambda,x,y,s,\alpha,\beta} \; & \{\lambda : Ax - \lambda\triangle b = b,\; x_{\mathcal{B}\cup\mathcal{T}} \in \mathcal{K}_{\mathcal{B}\cup\mathcal{T}},\; x_{\mathcal{N}} = 0,\; x_{\mathcal{R}} = \alpha x^*_{\mathcal{R}},\; \alpha \geq 0,\\ & A^T y + s - \lambda\triangle c = c,\; s_{\mathcal{N}\cup\mathcal{T}} \in \mathcal{K}_{\mathcal{N}\cup\mathcal{T}},\; s_{\mathcal{B}} = 0,\; s_{\mathcal{R}} = \beta s^*_{\mathcal{R}},\; \beta \geq 0\},\end{aligned}$$

where $\mathcal{K}_{\mathcal{B}\cup\mathcal{T}}$ is the Cartesian product of the cones $\mathcal{K}^i_q$ such that $i \in \mathcal{B}\cup\mathcal{T}$, $\mathcal{K}_{\mathcal{N}\cup\mathcal{T}}$ is defined analogously. Proof of this result for computing λ_ℓ and λ_u can be found in Theorem 4.1 in [153]. Alternatively, the constraints of the problems above can be completely rewritten in terms of the solution set instead of the index set, i.e., constraints $\{x_{\mathcal{B}\cup\mathcal{T}} \in \mathcal{K},\; x_{\mathcal{N}} = 0,\; x_{\mathcal{R}} = \alpha x^*_{\mathcal{R}},\; \alpha \geq 0\}$ can be written as $\{x \in \mathcal{K},\; x \circ s^* = 0\}$.

The optimization problems for computing the endpoints λ_ℓ and λ_u of the current invariancy interval are SOCO optimization problems due to the fact that

constraints of the form $x_{\mathcal{R}} = \alpha x^*_{\mathcal{R}}$, $\alpha \geq 0$ are linear (the invariancy interval can be a singleton, unlike in the QO case). In contrast, if we use the definition of the optimal partition π, constraints $x_{\mathcal{R}} \in \mathrm{bd}\,\mathcal{K}$ are non-linear and are not second-order cone representable.

The results obtained by Yildirim [153] for the simultaneous perturbation case in conic optimization and by using the geometric definition of the optimal partition are directly linked to our findings. In his paper, Yildirim proved that the optimal value function is quadratic on the current invariancy interval. Although Yildirim's and our results are very interesting in the light of extending the parametric optimization techniques to SOCO problems, the obstacles, discussed in the remaining of this section, prevent direct mapping of them to algorithm design and implementation.

Unlike for parametric LO and QO problems, the optimal partition π_r for SOCO may change continuously, that poses difficulties for identifying all invariancy intervals for parametric SOCO. For the intervals of the parameter λ, where the optimal partition π_r is not changing continuously, the optimal value function is quadratic (see Proposition 5.1 in [153]). Another way to say it, for parametric SOCO we can have a continuum of changing transition points until we find an invariancy interval. In general, the optimal value function is piecewise-quadratic and it is quadratic on every invariancy interval. For the intervals, where the optimal partition changes continuously, we obtain the regions of non-linearity of $\phi(\lambda)$ and there is no known way of describing $\phi(\lambda)$ completely on those intervals.

The intervals where the optimal partition π_r changes continuously, represent a curve on the boundary of the quadratic cone. Similarly, if the optimal partition is defined as π, the intervals with $\mathcal{R} \neq \emptyset$ represent a curve on the quadratic cone surface. Characterization of those curves and finding a computable description of them will allow identifying all invariancy intervals and computing the optimal value function. While those curves are conjectured to have hyperbolic shape, there no results characterizing those curves that we are aware of. To get a computational algorithm for parametric SOCO, this charac-

terization is a missing ingredient. Another remaining open problem is to find a rounding procedure for SOCO problems to identify exact optimal solutions.

Algorithms for computing the optimal value function $\phi(\lambda, \epsilon)$ for parametric SOCO problems are subject of future research as there are no algorithms for parametric SOCO optimization. Invariancy regions corresponding to the definition of the optimal partition π are illustrated by an example in Section 7.4. That example also highlights the difficulties that arise during bi-parametric SOCO.

Chapter 9

Conclusions and Future Directions

In this thesis we considered techniques for solving multiobjective optimization problems and their parametric counterparts. By formulating and solving *multi-objective optimization* problems as *parametric optimization* problems we bridged the gap between the two fields and unified the theory and practice of multiobjective and parametric optimization. Some classes of multiobjective optimization problems that include linear, convex quadratic and potentially second-order conic optimization problems can be efficiently solved using parametric optimization algorithms. In particular, parametric optimization techniques developed in this thesis give us a practical tool for solving multiobjective quadratic optimization problems. Parametric optimization allows not only computing Pareto fronts (efficient surfaces), but also identifying piece-wise structure of those surfaces. Structural description of Pareto fronts gives functional form of each of its pieces and thus helps decision makers to make better decisions.

We extended the existing theory of parametric quadratic optimization to simultaneous perturbation sensitivity analysis when the variation occurs in both the right-hand side vector of the constraints and the coefficient vector of the linear term of the objective function. First, in our analysis the rate of variation, represented by the parameter λ, was identical for both perturbation vectors $\triangle b$ and $\triangle c$, which allows solving simultaneous perturbation *uni-parametric quadratic*

optimization problems. One of the main results is that the invariancy intervals and transition points can be determined by solving auxiliary linear or quadratic problems. As we already mentioned, all auxiliary problems can be solved in polynomial time.

We have extended the uni-parametric simultaneous perturbation results for quadratic optimization to the bi-parametric case. By solving a series of uni-parametric QO problems, the algorithm outlined in the thesis allows identifying all invariancy regions where the optimal partition is invariant. On each invariancy region we can also compute the optimal value function and maximally complementary solutions. To the best of our knowledge, this thesis and paper [57] are the first attempts to study systematically *bi-parametric convex quadratic optimization* problems with different parameters in the coefficients of the objective function and the right-hand-side of the constraints.

Finally, we developed and implemented the *algorithm* that represents a sequence of linear and quadratic auxiliary problems to identify all invariancy regions and graph the optimal value function. Even though all presented auxiliary optimization problems can be solved in polynomial time by IPMs and the number of different optimal partitions is finite, enumeration of all invariancy regions may not be achieved in polynomial time due to the fact that the number of different optimal partitions may increase exponentially with the cardinality of the index set. That is why the algorithm presented is linear in the output size, but not in the input size.

The implementation of the algorithm for parametric optimization and testing results suggest that our implementation is well suitable for solving small- and medium-size parametric quadratic problems and can be used for large-scale problems as well with some caution connected to the precision of solutions produced by a quadratic solver. We illustrated our implementation on a set of simple parametric quadratic problems as well as on a number of selected *applications.*

As it was mentioned in the introduction, the most famous application of the convex QO multiobjective analysis is the mean-variance portfolio optimization

problem introduced by Markowitz [95]. Our method allows us to analyze not only the original Markowitz model, but also some of its extensions. Extensions include minimizing linear transaction cost in addition to optimizing the mean and the variance, computing sparse and stable Markowitz frontiers and analyzing multiobjective robust mean-variance optimization formulations. Other *practical problems*, where our methodology can be used, are cancer treatment planning and multi-drug composition-activity analysis among others. We illustrated solving all these multiobjective optimization *case studies* with our parametric algorithms.

In addition to developing and applying parametric optimization techniques, we studied two novel financial optimization frameworks, namely portfolio *credit risk optimization* and *portfolio replication.*

We analyzed different approaches to the portfolio *credit risk optimization* problem within a conditional independence framework and compared optimization of quantile-based risk measures with the traditional Markowitz mean-variance optimization formulation. Optimization of quantile-based risk measures using the MC approximation, requires a large number of scenarios. If the number of instruments in the portfolio is relatively large, an efficient LLN or especially CLT approximation of the conditional distribution of portfolio losses can be utilized in the conditional independence framework. These two approximations can be used to obtain a significant reduction in the number of scenarios required for portfolio optimization. After comparing different credit-risk optimization formulations, the one that we recommend for practical implementation, is the nonlinear CLT approximation. The CLT approximation is attractive for optimization due to:

- Producing consistently better results than MC sampling with only 10% of the data;
- Acceptable performance when solving the nonlinear formulation;
- Being relatively robust to violations of the portfolio granularity condition.

The MC linear formulation is competitive with the CLT formulation only if the number of groups is very small (less than 10) and if groups consist of a very small

number of counterparties. Mean-variance may not always produce good results and is quantile-independent.

In this thesis we also analyzed existing, and developed new algorithmic tools for the *portfolio replication* problem, as well as tested these methodologies on real large-scale data. As the goal was to develop practical multiobjective optimization models for portfolio replication, we have looked at different formulations based on regularized optimization. We worked on identifying which assets should be included in the tradable universe for the replicating portfolios, and developed a procedure for removing duplicate and highly correlated assets. Special attention was paid to research issues related to replication overfitting and minimizing different objective functions, as well as other characteristics to improve replication. Reduction in replication overfitting was achieved by jointly minimizing portfolio replication error and trading costs with different weights that represent the relative importance of each objectives. This problem is solved as a multiobjective optimization problem. We demonstrated the benefits of the multiobjective optimization formulation, and computed efficient frontiers that are used to justify and explain the usage of trading penalties in problem formulation.

There are many possible extensions and future work in both theoretical and implementation directions. Some of them are discussed in the sequel.

(i) Preliminary results on parametric second-order cone optimization are described in Chapter 8. The first priority here would be to extend the methodology and develop a practical algorithm for solving uni- and bi-parametric second-order cone optimization problems. It is also worthwhile exploring parametric SOCO applications to financial models.

(ii) Integration of parametric optimization techniques that use optimal bases, optimal set invariancy and optimal partition invariancy into a unified framework remains to be done. There are many publications that address different aspects of parametric optimization, but there is no study that puts those techniques together and describes how well those perform for dif-

ferent classes of optimization problems. Additional work has to be done on classifying multiobjective optimization problems for which the Pareto efficient frontier has identifiable structure.

(iii) Future work on portfolio credit risk optimization includes going from single-period to multi-period setting, incorporating stochastic exposures and stochastic recoveries into formulations and efficiently integrating credit derivatives into an optimization framework. From the methodological side, we would look at importance sampling and worst-case (robust) optimization in the credit risk setting. Finally, data and formulation features that affect optimization performance need to be studied.

(iv) Most portfolio replication studies tend to consider only a simple type of trading constraint, which assigns an identical trading cost to all instruments. More sophisticated costing schemes can potentially yield better results. It is worthwhile investigating which trading costs are likely to produce the best replicating portfolio. Addressing this question includes evaluating a number of alternative costing schemes based on their out-of-sample performance, comparing performance of trading cost restrictions to that of cardinality constraints, and assessing the impact of optimization formulations, e.g., minimizing absolute vs. squared cash flow deviations, level of cash flow bucketing, etc., on the effectiveness of costing schemes.

From the computational/implementation side we suggest the following future directions:

(i) Make the implementation of parametric optimization algorithms more robust and improve its stability with respect to numerical issues. For instance, developing and implementing algorithms for optimal partition identification requires dealing with a number of theoretical and computational problems. Recovery from numerical issues in computing boundaries of invariancy regions involve careful implementation of safeguards against imprecise solutions of auxiliary problems.

(ii) Extend the implementation to multi-parametric QO case. One of the challenges for multi-parametric QO optimization is computing and storing data structures describing complex and multidimensional geometric objects.

(iii) Extend the implementation to parametric SOCO case. Parametric SOCO case still requires some experimentation to get closer to practical implementation.

(iv) Implementing parametric optimization into optimization software packages remains one of the challenges. Unfortunately, available software for parametric optimization is very limited. Commercial optimization packages such as CPLEX and MOSEK include optimal basis based sensitivity analysis for LO. MOSEK is the only package that provides optimal partition based sensitivity analysis for LO as an experimental feature.

(v) Experiment with warm-start strategies in the context of parametric optimization. Parametric optimization algorithms need to utilize existing and new warm-start strategies to speed-up solutions of auxiliary problems, that are near-identical in terms of problem data.

Appendix A

Credit Risk Optimization: Computational Tests

This appendix contains the description of the computational tests performed for portfolio credit risk optimization formulations described in Section 6.3.

Computational Framework and Performance Metrics

In our computational tests, the set of feasible solutions Ω is given by the constraints

- the value of the portfolio remains at its initial value (v_P);
- the expected return of the portfolio is at least that of the initial portfolio (r_P);
- the group positions have fixed lower and upper bounds ($\underline{x}_i \leq \overline{x}_i$, for the i-th group).

This leads to the following optimization problem:

$$\begin{aligned} \min_{x} \quad & g\left[\mathcal{L}(x)\right] \\ \text{s.t.} \quad & \sum_{i=1}^{N_{\mathrm{G}}} v_i x_i = v_P, \\ & \sum_{i=1}^{N_{\mathrm{G}}} r_i v_i x_i \geq r_P \cdot v_P, \\ & \underline{x}_i \leq x_i \leq \overline{x}_i, \ i = 1, \ldots, N_{\mathrm{G}}, \end{aligned}$$

Other types of constraints are possible, such as a total budget constraint (e.g., trading budget of 5% of portfolio value, including transaction costs); or limits on the total positions in certain categories, e.g., a particular credit rating or geographic region.

For the LLN, CLT and MC formulations, there are two sources of error affecting the quality of an optimal solution: (i) systemic sampling error due to the choice of M; and (ii) modeling (or formulation) error due to approximating the conditional loss distributions. (For the MC formulation, the choice of K is the influential parameter.)

To isolate and quantify these errors, we measure:

- *Out-of-sample actual risk*: the risk measure for the optimal portfolio under a given formulation, evaluated by MC with a large number of out-of-sample

scenarios. The out-of-sample actual risk is assumed therefore to be based on the true loss distribution.

- *In-sample actual risk*: the risk measure for the optimal portfolio under a given formulation, evaluated by MC with the same systemic scenarios that are used for optimization and a large number of idiosyncratic samples for each systemic sample. *In-sample actual risk* is assumed therefore to be based on the true conditional loss distributions. The approximation error due to systemic sampling, is thus isolated and gauged by the difference between *out-of-sample actual risk* and *in-sample actual risk*.
- *In-sample reported risk*: the objective function value reported for the risk measure for the optimal portfolio under a given formulation. The approximation error due to the choice of optimization formulation (equivalently, the choice of approximation in the conditional loss distributions) or model error, is thus isolated and gauged by the difference between *in-sample actual risk* and *in-sample reported risk*. The difference between *out-of-sample actual risk* and *in-sample reported risk*, contains the combined effect of systemic sampling error and formulation approximation error.

Note that for the variance and second-moment formulations, only out-of-sample actual risk is applicable.

Unless specific mention is made otherwise, the optimization problems were run with the following default settings:

- positions are between zero and twice the initial position: $0 \leq x_i \leq 2x_i^0$;
- $M = 10,000$ optimization scenarios (systemic);
- VaR and ES are calculated for the 99th percentile; i.e., $\alpha = 0.999$;
- the results are the averages over 5 trial runs (different sets of systemic scenarios are used for each trial).

In our computational experiments, we test different group sizes and group compositions. This allows us to make conclusions about the effect of the group sizes, etc.

We have used the following optimization packages for implementing our formulations:

- IPOPT – nonlinear solver [143];
- CPLEX – LO, QO, SOCO solver (version 11.2) [74];
- MOSEK – LO, QO, SOCO, convex nonlinear solver (version 5) [105].

All of these optimization packages can be called directly both from C/C++ and from the MATLAB environment. This feature and the efficiency of optimization algorithms implemented in those solvers make them efficient for both prototyping and high-performance practical implementation.

The formulations described in the previous sections were tested and compared using two portfolios. The first portfolio is adapted from [127]. It consists of 3000 counterparties depending on 50 credit drivers and graded with 8 credit states (credit-state migrations). The second one, that is a proprietary portfolio, consists of 7470 counterparties with 34 credit drivers and only 2 credit states that may be default or nondefault. All the formulations are implemented with the assumption that the recovery rates are deterministic.

Results: 3000 Counterparties and Credit-State Migrations

Credit drivers in the portfolio are industry/country indices and each counterparty depends on one credit driver with $0.42 \leq \beta \leq 0.65$.

Grouping is done at random with four grouping schemas: 10 groups of 300 CPs, 50 groups of 60 CPs, 300 groups of 10 CPs, and 3000 groups of 1 CP.

For *out-of-sample actual risk*, $M = 6,000,000$ and $K = 1$. For *in-sample actual risk*, $M = 10,000$ and $K = 150$. For *in-sample reported risk*, $M = 10,000$ and, for MC formulation MC(1), $K = 1$, while for MC formulation MC(20), $K = 20$.

The out-of-sample evaluation of the optimization results are presented in Figure A.1 for VaR and in Figure A.2 for ES. These figures show reductions in

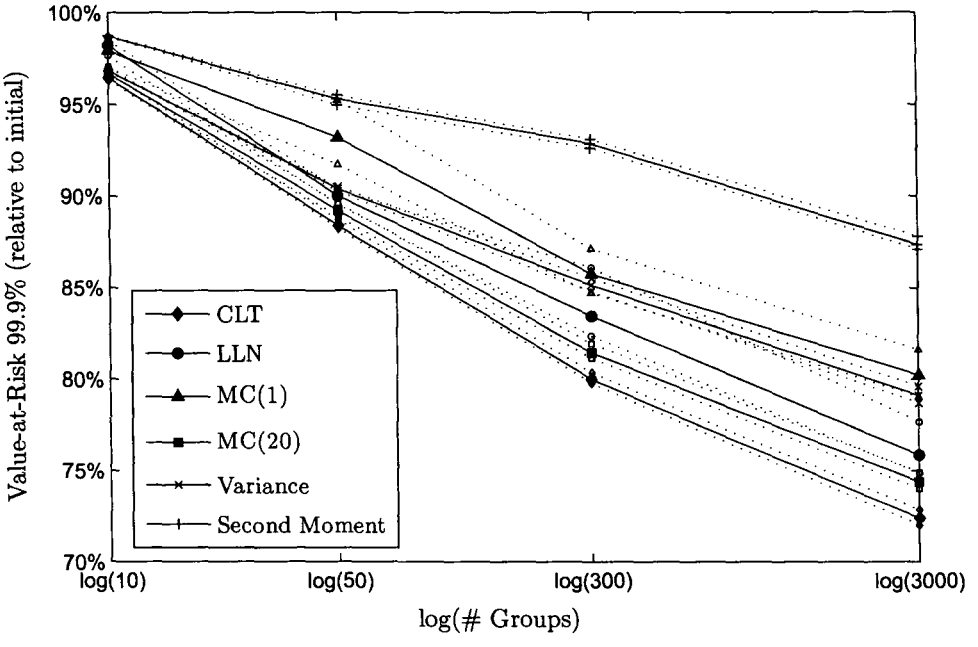


Figure A.1: Out-of-Sample VaR.

VaR and ES relative to the initial portfolio (100%). The horizontal axes display the group sizes in a logarithmic scale. The solid lines are the averages over 5 trials and the dotted lines are the best and the worst results for the same 5 trials. As out-of-sample testing incorporates both systemic sampling error and formulation error, we can make conclusions about the overall performance of the optimizations. The CLT formulation always has the best performance followed by MC sampling with $K = 20$ idiosyncratic samples per systemic scenario, which in turn is much better than MC sampling with $K = 1$. Extrapolating, we expect that MC sampling with very large K would match and possibly exceed the performance of CLT. However, increasing K leads to a larger, more computationally demanding problem.

There is little difference between the formulations' results when the number of groups is small, because there is less opportunity to restructure the portfolio. The differences become more pronounced as more trading flexibility, i.e., smaller

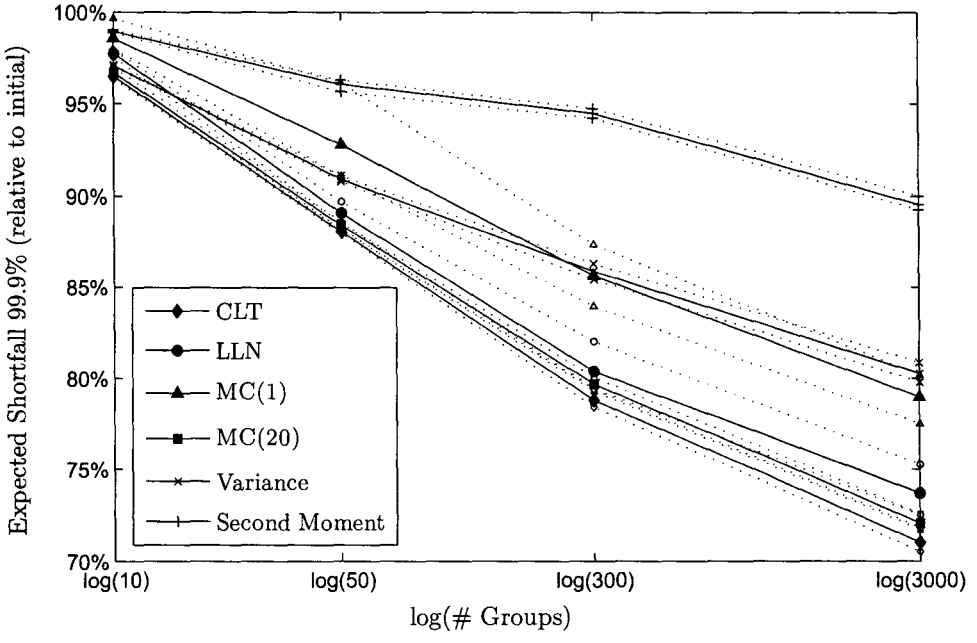


Figure A.2: Out-of-Sample Expected Shortfall.

groups, is introduced. Formulations that show relatively good performance for all grouping schemas are the CLT, MC(20) and LLN. The LLN formulation performance is acceptable due to fine granularity of the optimal portfolios (narrow budget preserves the fine granularity of the initial portfolio). We investigate the issue of granularity in more detail later in this appendix. CLT, MC(20) and LLN formulations are quite separated from each other with minimal overlaps when we look at the best and the worst performance over 5 trials. It is more pronounced for the VaR models, partially due to the heuristic algorithm that we utilize for the MC and LLN approximations. This also explains why the performance of the MC and LLN formulations is not as good when minimizing VaR, as compared to the ES minimization.

The next step in evaluating the computational performance, is to look at the approximation quality of optimization formulations, which is summarized in Figure A.3 for VaR, and Figure A.4 for ES. These figures display the out-of-

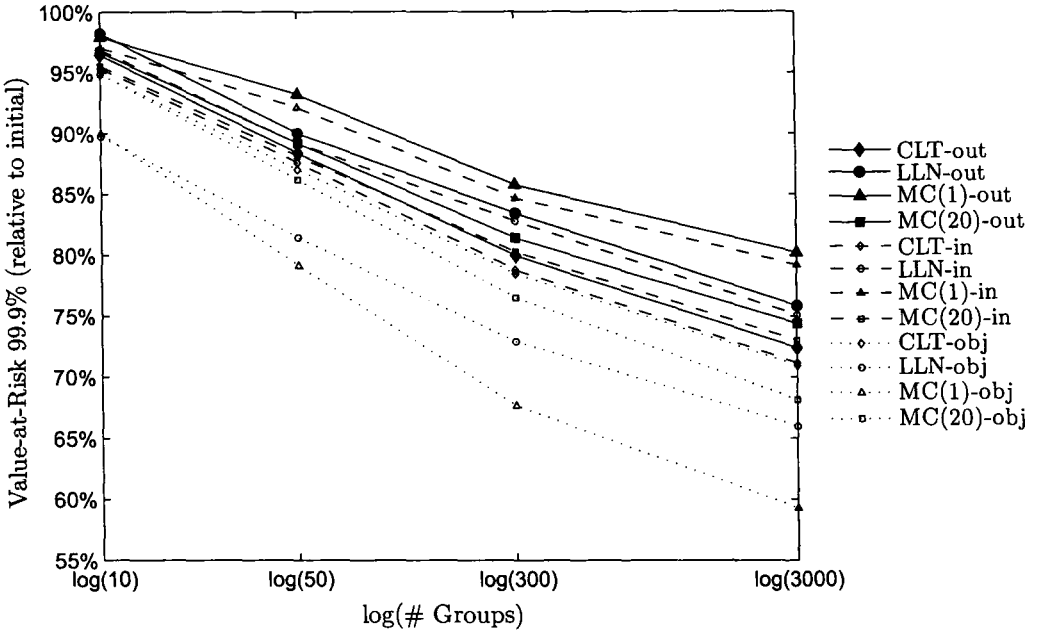


Figure A.3: Approximation Quality for the VaR Optimization Problems.

sample actual risk, and the in-sample actual and reported risks for the various formulations.

The systemic sampling error is about the same for all formulations, as they all use the same number of systemic scenarios, M. The CLT approximation has the smallest formulation error, while the formulation error of MC(1) is the largest. For MC, the formulation error decreases when group-size increases. This observation can be explained by the way in which the tail of the conditional group-loss distribution is represented in each of the formulations. For MC sampling with only a few specific scenarios, the tail may be severely underestimated if the conditional CP default probabilities are very small. The optimizer will take advantage of this possibility by overinvesting in groups which, by chance, experience a small number of defaults, even though they may be more riskier theoretically. Thus, for a large number of groups, that in our experiment imply a small number of CPs per group, the number of these chance opportunities

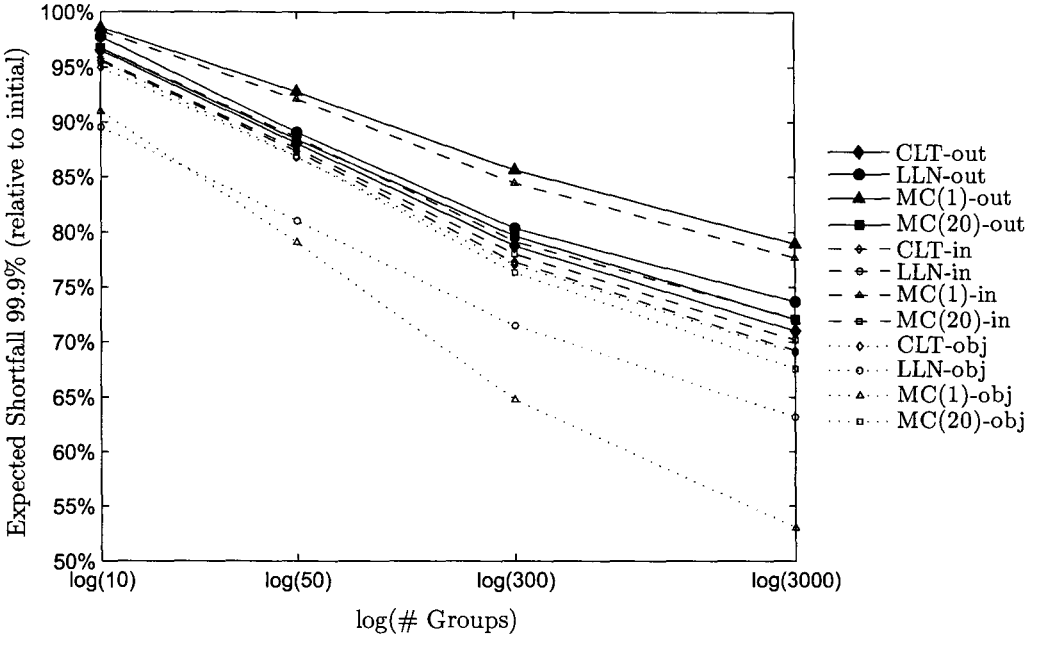


Figure A.4: Approximation Quality for the Expected Shortfall Optimization Problems.

will tend to be large, which leads to high formulation error. In contrast, the LLN and CLT formulations do not provide the optimizer such opportunities, as the conditional group-loss distributions are derived analytically rather than empirically.

It is known that the quality of the LLN and CLT approximations deteriorates as the portfolio becomes more granular, i.e., exposures are increasingly concentrated in a small number of dominant groups. To test granularity effects, for this test we modified the initial portfolio to be as finely grained as possible. Further, the original portfolio value v_P is distributed equally among all CPs in the portfolio. The CPs were left in the original grouping scheme for 50 equally sized groups and the optimization problems were solved with five sets of progressively wider trading limits: [1, 1], [0, 2], [0, 15], [0, 30] and [0, 50]. As the trading limits become larger, there is a greater chance for obtaining granular portfolios

and thus, more potential for poor results from the LLN and CLT approximations. As a granularity measure, we use the *Herfindahl-Hirschman Index* or HHI which is the sum of squared group-weights[1]. In our case, the initial HHI equals 0.02, which is the lowest possible value for a portfolio comprising 50 groups.

Figure A.5 plots the formulation error, i.e., the relative difference between in-sample actual risk and in-sample reported risk, against granularity for the optimal portfolios obtained by the LLN, CLT, MC(1) and MC(20) formulations. The results suggest that the LLN approximation is far more susceptible to the adverse effects of granularity than the CLT formulation. While wider trading limits tend to produce more granular portfolios in general, the optimal MC and CLT portfolios maintain some degree of diversification for which the HHI does not exceed 0.25. In contrast, LLN portfolios become progressively more concentrated and exhibit increasingly large formulation errors.[2] This is explained by the fact that, to minimize VaR or ES, the LLN formulation tries to reduce only the portfolio's conditional expected losses while the CLT formulation considers both the conditional expectation and the conditional variance. Reducing the variance of the portfolio losses entails smaller positions, i.e., greater diversification. It was also observed that the relative rankings of the sampled group losses across the systemic scenarios were much more variable than those of the expected group losses.[3] The greater degree of diversification present in the optimal MC portfolios relative to those of LLN is a reflection of this fact.

To see the effect of increasing the number of systemic samples for use in optimization, we ran our formulations with $M = 10,000$ and $M = 50,000$ systemic scenarios. For the MC approximation MC(20), in the case with $M = 10,000$ we use $K = 20$ and $K = 4$ for MC(4) in the case with $M = 50,000$,

[1]The group-weight for G_i, is $x_i v_i / V$. By definition, the HHI converges to zero for infinitely finely grained portfolios when the number of groups grows indefinitely.

[2]The large formulation errors associated with MC(1) are due more to the relatively small sample size than to granularity effects.

[3]While the loss of a particular group might be among the smallest losses sampled in one scenario and among the largest losses sampled in another, its expected loss tends to rank at approximately the same level in all scenarios.

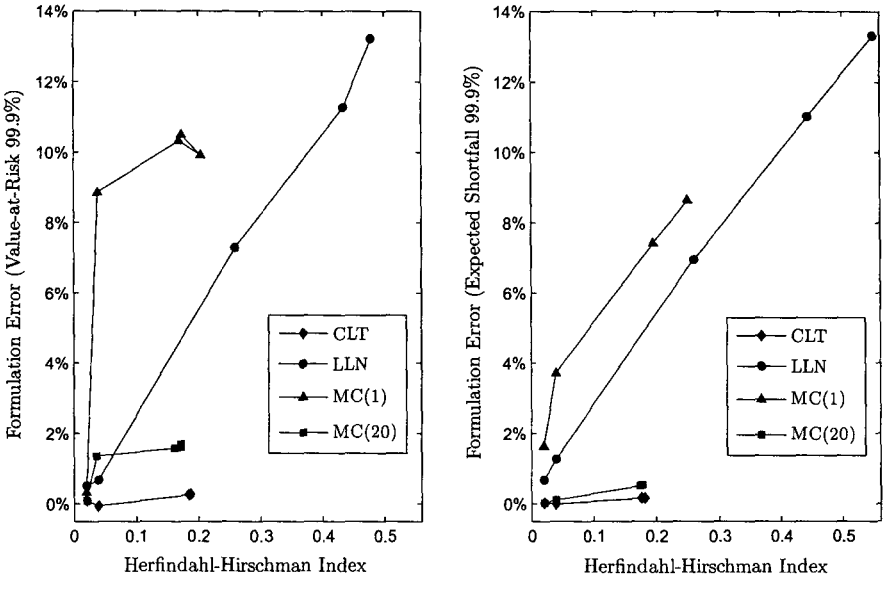


Figure A.5: Granularity Effect.

resulting in a total of 200,000 MC scenarios. Table A.1 describes the systemic sampling effect and shows how out-of-sample performance is affected by the number of systemic samples. Increasing the number of systemic samples by a factor of 5 produces slight improvement for the MC(1) formulation, and has a negligible effect for the others. It also turns out that CLT with $M = 10,000$ systemic samples, performs better than the other formulations with $M = 50,000$ systemic samples.

Analyzing the optimal group-weights for different random samples, makes it possible to compare the stability of formulations. Figure A.6 shows the ranges of optimal group-weights for 5 trials when minimizing ES, where for each trial different sets of (systemic) scenarios were used for optimization. Similarly to the sensitivity of the out-of-sample objective function value in Figure A.2, stable group-weights are produced by the CLT formulation, while the MC formulations require a larger number of scenarios to reach stability of the weights. For the mean-variance formulations, the variance-covariance matrix of losses and the vec-

Table A.1: Systemic Sampling Effect.

$VaR_{99.9\%}$	10,000 Systemic Samples			$VaR_{99.9\%}$	50,000 Systemic Samples		
	10 Groups	50 Groups	300 Groups		10 Groups	50 Groups	300 Groups
CLT	96.5%	88.4%	80.0%	CLT	96.3%	88.3%	79.6%
MC(20)	96.7%	89.2%	81.4%	MC(4)	96.6%	89.2%	81.3%
LLN	98.2%	90.0%	83.4%	LLN	97.4%	89.5%	82.1%
MC(1)	97.9%	93.2%	85.8%	MC(1)	97.1%	90.4%	82.9%

$ES_{99.9\%}$	10,000 Systemic Samples			$ES_{99.9\%}$	50,000 Systemic Samples		
	10 Groups	50 Groups	300 Groups		10 Groups	50 Groups	300 Groups
CLT	96.5%	88.1%	78.8%	CLT	96.4%	87.9%	78.4%
MC(20)	96.7%	88.4%	79.7%	MC(4)	96.6%	88.5%	79.3%
LLN	97.8%	89.1%	80.4%	LLN	97.6%	89.4%	79.7%
MC(1)	98.6%	92.8%	85.6%	MC(1)	96.9%	89.5%	80.8%

tor of mean losses are estimated from the $M = 10,000$ with $K = 150$ samples (1.5 million in-sample scenarios) and, consequently, are quite stable. This explains why mean-variance formulations produce the most stable weights among all the formulations that were considered.

We also evaluated the sensitivity of ES to the quantile level, see Figure A.7. We used 300 groups and $M = 50,000$ for optimization. For lower quantiles (95% or less), the performance of all formulations is very similar. For high quantiles (99.97%), the performance of CLT becomes more distinguishable from the other formulations. As Figure A.7 demonstrates, the CLT formulation has a 41% improvement over variance minimization, when the optimization results are evaluated for ES out-of-sample for the 99.97% quantile, but the variance minimization result is within 2% of the CLT formulation for the 95% quantile. It is known that mean-variance formulations perform well when the quantile moves towards the center of a distribution, as the Normal approximation becomes more appropriate. MC sampling formulations require more scenarios to capture losses in the extreme tail of the distribution and, as a result, they perform less effectively than for lower quantiles. The same reasoning applies for the LLN formulation which better approximates the tail for high quantiles than MC formulations for the sample sizes used in our tests; but as we already know, the LLN formulation

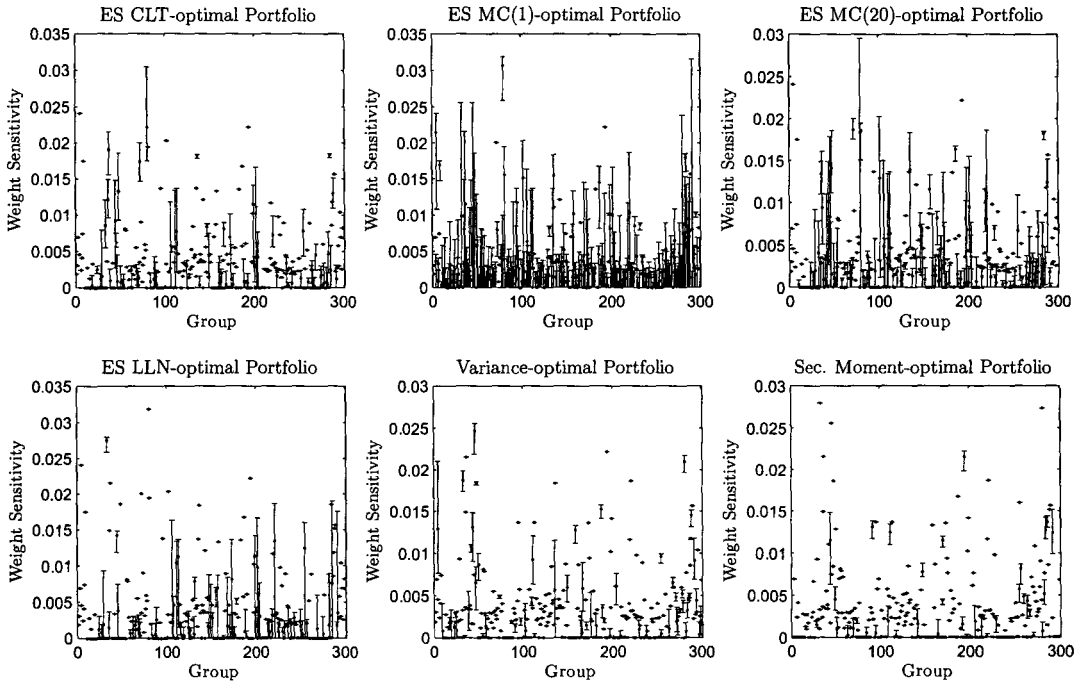


Figure A.6: Sensitivity of Group-Weights to Different Sets of Optimization Scenarios.

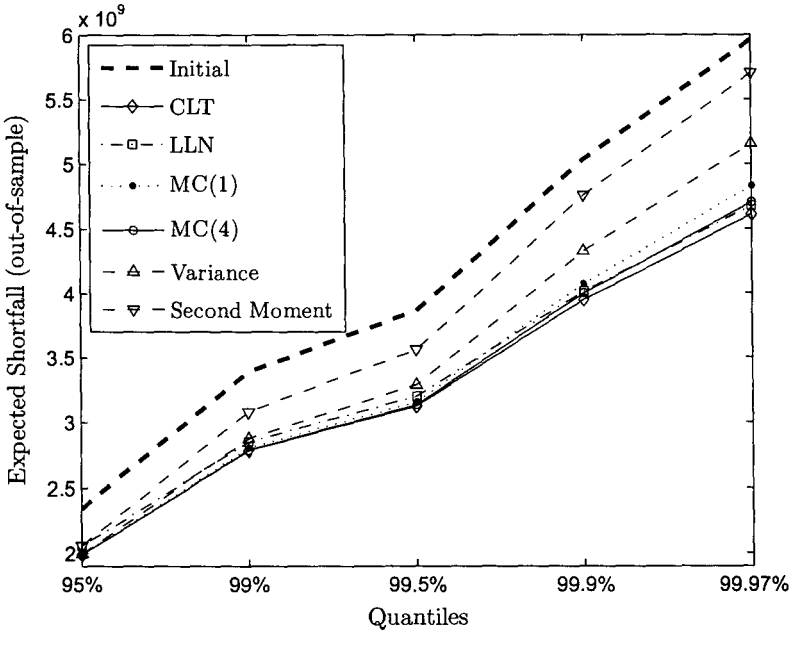


Figure A.7: Formulation Performances for Different Quantiles.

is very sensitive to the granularity effect. Note that as the tests in Figure A.7 are performed for a narrow budget, the granularity effect is limited.

Finally, we tested the performance of the formulations when the portfolio comprises a very small number of counterparties. This case is expected to be the worst for the CLT and LLN approximations, as the portfolio is no longer large and some counterparties can dominate within it. Sensitivity of the out-of-sample optimal solution to different sets of optimization scenarios is shown in Figure A.8. The results are presented for 10 and 20 groups with 1 CP per group. We can conclude that LLN fails for this setup. The MC(1) formulation has the highest sensitivity and produces unacceptable results on average. CLT becomes unreliable for portfolios with a small number of counterparties. MC sampling MC(20) always works, but requires a large number of idiosyncratic scenarios per systemic scenario. It is necessary to point out that this setting is unrealistic from the practical point of view, as real credit portfolios cannot be that small.

One observation from our computational experiments is that variance minimization in the mean-variance context tends to improve the quantile-based measures, see Figures A.1, A.2 and A.7. One possible reason is the fact that counterparties with high probability of default have high variance of monetary losses. This is due to the fact that we use a Bernoulli mixture model [130] for defaults and migrations, and consequently, variance minimization leads to reduction of the quantile risk measures, as the counterparties with high default probabilities are removed from the optimal portfolio.

Table A.2 summarizes running times for different optimization formulations. Evidently, solving the nonlinear problems for CLT is competitive with solving the linear problems in the MC-sampling formulations, at a comparable level of (out-of-sample) accuracy. The computational tests were run on a server with 8 x Opteron 885 CPU, 16 cores, 64 GB RAM, each jobs run on 1 core. The VaR optimization for MC(20) was run in parallel mode on 4 threads.

We also ran computational tests with more realistic counterparty groupings. Counterparties depending on the same credit driver and which are in the

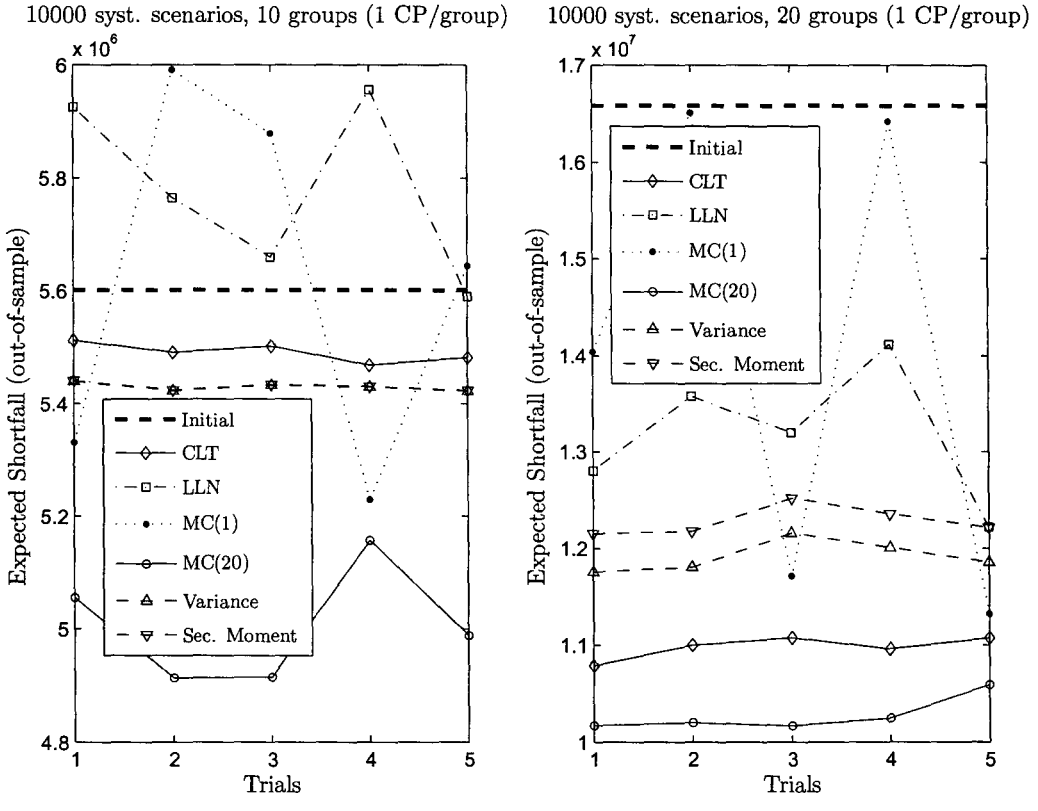


Figure A.8: Effect of Small Number of Groups for Expected Shortfall Optimization.

Table A.2: Optimization Problem Performance: Elapsed Time in Seconds.

Model	Solver	$\mathbf{ES}_{99.9\%}$			
		10 Groups	50 Groups	300 Groups	3000 Groups
CLT	IPOPT	4–8	6–8	14–83	81–1090
LLN	CPLEX	1	1–2	6–8	73–86
MC(1)	CPLEX	1	1–2	6–10	14–115
MC(20)	CPLEX	137–155	233–279	461–578	1050–1280
Model	Solver	$\mathbf{VaR}_{99.9\%}$			
		10 Groups	50 Groups	300 Groups	3000 Groups
CLT	IPOPT	4–25	5–7	14–55	400–1643
LLN	CPLEX	2–3	6–8	46–50	791–1025
MC(1)	CPLEX	2–3	8–10	46–69	2436–3312
MC(20)	CPLEX	3620–4080	2382–2777	6522–8563	39273–86383
Model	Solver	**Variance**			
		10 Groups	50 Groups	300 Groups	3000 Groups
Uncond.	MOSEK	<1	<1	1	682–719

same credit state at the beginning of the time period belong to the same group, resulting in 301 groups with a variable number of counterparties in each of them. Table A.3 summarizes out-of-sample performance relative to the performance of the initial portfolio of the optimization formulations for $0 \leq x_i \leq 2x_i^0$, and $0.8x_i^0 \leq x_i \leq 1.2x_i^0$ trading limits. The narrow trading limits are more realistic when practical portfolio rebalancing is performed. The CLT formulation exhibits the best performance for both trading limits.

Results: 7470 Counterparties and Default/Nondefault Credit States

To confirm our findings while working with the portfolio of 3000 counterparties and 8 credit states, we ran some of the computational tests for the portfolio with 7470 counterparties and only 2 credit states, default and nondefault. In both portfolios the credit drivers are the industry/country indices, and each counterparty depends on one credit driver, but $0.30 \leq \beta \leq 0.95$ for the second

Table A.3: Out-of-sample Performance for the Portfolio with 3000 CPs.

Model	Risk Measure	10000 Systemic Scenarios, 301 Groups	
		Trading Limits [0.8, 1.2]	Trading Limits [0, 2]
CLT	$ES_{99.9\%}$	93.62%	73.19%
	$VaR_{99.9\%}$	93.92%	74.53%
MC(20)	$ES_{99.9\%}$	93.63%	73.47%
	$VaR_{99.9\%}$	94.78%	75.31%
LLN	$ES_{99.9\%}$	93.69%	74.98%
	$VaR_{99.9\%}$	95.00%	77.29%
MC(1)	$ES_{99.9\%}$	94.94%	78.77%
	$VaR_{99.9\%}$	96.50%	79.86%
Variance	$ES_{99.9\%}$	95.92%	83.38%
	$VaR_{99.9\%}$	95.64%	82.05%
Second Moment	$ES_{99.9\%}$	97.89%	91.61%
	$VaR_{99.9\%}$	97.53%	89.55%

Table A.4: Out-of-sample Performance for the Portfolio with 7470 CPs.

Model	Risk Measure	10000 Systemic Scenarios, 277 Groups	
		Trading Limits [0.8, 1.2]	Trading Limits [0, 2]
CLT	$ES_{99.9\%}$	89.84%	52.67%
	$VaR_{99.9\%}$	90.27%	54.40%
MC(20)	$ES_{99.9\%}$	89.84%	52.89%
	$VaR_{99.9\%}$	90.86%	55.15%
LLN	$ES_{99.9\%}$	89.87%	52.87%
	$VaR_{99.9\%}$	91.37%	55.64%
MC(1)	$ES_{99.9\%}$	90.39%	55.31%
	$VaR_{99.9\%}$	93.12%	58.61%
Variance	$ES_{99.9\%}$	90.01%	54.16%
	$VaR_{99.9\%}$	90.30%	55.20%
Second Moment	$ES_{99.9\%}$	91.26%	59.61%
	$VaR_{99.9\%}$	91.50%	60.19%

portfolio. Similar to the results in Table A.3, we used realistic counterparty groupings, based on common credit driver and initial credit state, resulting in 277 groups with a variable number of counterparties in each of them. Table A.4 summarizes the out-of-sample performance of the optimization formulations for this portfolio.

Table A.5: Out-of-sample Performance for the Portfolio with 3000 CPs by Including Default/Nondefault Credit States Only.

Model	Risk Measure	10000 Systemic Scenarios, 301 Groups, Trading Limits [0, 2]	
		Migrations	Default/Nondefault
CLT	$ES_{99.9\%}$	73.19%	72.45%
	$VaR_{99.9\%}$	74.53%	74.88%
MC(20)	$ES_{99.9\%}$	73.47%	72.67%
	$VaR_{99.9\%}$	75.31%	75.23%
LLN	$ES_{99.9\%}$	74.98%	75.54%
	$VaR_{99.9\%}$	77.29%	79.00%
MC(1)	$ES_{99.9\%}$	78.77%	79.57%
	$VaR_{99.9\%}$	79.86%	79.88%
Variance	$ES_{99.9\%}$	84.08%	84.08%
	$VaR_{99.9\%}$	82.05%	82.14%
Second Moment	$ES_{99.9\%}$	91.61%	96.12%
	$VaR_{99.9\%}$	89.55%	93.37%

The out-of-sample performance of optimization formulations for [0.8, 1.2] and [0, 2] trading limits confirms the same findings as before: the CLT formulation outperforms all the other optimization formulations when the number of counterparties in a portfolio is relatively large. Our numerical experiments show that allowing credit-state migrations, or just default/nondefault credit states, does not have any observable effect on performance of the optimization formulations other than that the discrepancy between the formulations gets smaller. Note that optimization results from the mean-variance formulation are much closer to those of other formulations. This raises the question of whether the "convergence" of the different formulations is due to a feature of the model – default/nondefault vs. migration – or due to some feature of the initial portfolio. We ran the 3000 CP portfolio in default/nondefault mode and found that the differences between formulations were sustained as it can be seen in Table A.5. Therefore, we conclude that the convergence is due to some features of the portfolios themselves, rather than using a migration model vs. a default only model. Investigating which portfolio features, e.g., sensitivities of counterparties to credit drivers or portfolio size, may deteriorate or improve the moment-based formulation performance is a subject of future research.

Bibliography

[1] I. Adler and R. D. C. Monteiro, *A geometric view of parametric linear programming*, Algorithmica, 8 (1992), pp. 161–176.

[2] S. Alexander, T. F. Coleman, and Y. Li, *Minimizing CVaR and VaR for a portfolio of derivatives*, Journal of Banking and Finance, 30 (2006), pp. 583–605.

[3] N. M. Alexandrov and M. Y. Hussaini, eds., *Multidisciplinary Design Optimization: State of the Art*, Proceedings in Applied Mathematics Series, No. 80, SIAM, Philadelphia, PA, USA, 1997.

[4] F. Alizadeh and D. Goldfarb, *Second-order cone programming*, Mathematical Programming, Ser. B, 95 (2003), pp. 3–51.

[5] F. Andersson, H. Mausser, D. Rosen, and S. Uryasev, *Credit risk optimization with Conditional Value-at-Risk criterion*, Mathematical Programming, Ser. B, 89 (2001), pp. 273–291.

[6] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer, *Non-Linear Parametric Optimization*, Birkhäuser Verlag, Basel, Switzerland, 1983.

[7] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, USA, 2009.

[8] A. BEN-TAL AND A. NEMIROVSKI, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, MPS-SIAM Series on Optimization, MPS/SIAM, Philadephia, PA, USA, 2001.

[9] R. A. BERK, *Statistical Learning from a Regression Perspective*, Springer Series in Statistics, Springer, New York, NY, USA, 2008.

[10] A. B. BERKELAAR, B. JANSEN, C. ROOS, AND T. TERLAKY, *An interior-point approach to parametric convex quadratic programming*, Working Paper, Erasmus University Rotterdam, Rotterdam, The Netherlands, 1997.

[11] A. B. BERKELAAR, B. JANSEN, K. ROOS, AND T. TERLAKY, *Basis- and partition identification for quadratic programming and linear complementarity problems*, Mathematical Programming, 86 (1999), pp. 261–282.

[12] A. B. BERKELAAR, C. ROOS, AND T. TERLAKY, *The optimal set and optimal partition approach to linear and quadratic programming*, in Advances in Sensitivity Analysis and Parametric Programming, T. Gal and H. J. Greenberg, eds., Kluwer Academic Publishers, Boston, MA, USA, 1997, chapter 6, pp. 6-1 – 6-44.

[13] J. F. BONNANS AND H. RAMÍREZ C., *Perturbation analysis of second-order cone programming problems*, Mathematical Programming, Ser. B, 104 (2005), pp. 205–227.

[14] C. BOSE, J. GONG, AND O. ROMANKO, *Optimization of multi-drug composition for the most efficacious action*, in Proceedings of 11th PIMS Industrial Problem Solving Workshop, Edmonton, Canada, June 2007, PIMS. Forthcoming.

[15] S. P. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[16] J. BRODIE, I. DAUBECHIES, C. D. MOL, D. GIANNONE, AND I. LORIS, *Sparse and stable Markowitz portfolios*, Proceedings of the National Academy of Sciences of the United States of America, 106 (2009), pp. 12267–12272.

[17] C. BURMEISTER AND H. MAUSSER, *Using trading restrictions in replicating portfolios*, Life & Pensions, November (2009), pp. 36–40.

[18] C. BURMEISTER, H. MAUSSER, AND R. MENDOZA, *Techniques for managing tracking error*, in Portfolio Analysis: Advanced Topics in Performance Measurement, Risk and Attribution, T. P. Ryan, ed., Risk Books, London, UK, 2006, chapter 12, pp. 263–281.

[19] C. BURMEISTER, H. MAUSSER, AND O. ROMANKO, *Using trading costs to construct better replicating portfolios*, in Proceedings of the Enterprise Risk Management Symposium, Chicago, IL, USA, April 12-14, 2010.

[20] E. J. CANDÈS, M. B. WAKIN, AND S. P. BOYD, *Enhancing sparsity by reweighted ℓ_1 minimization*, Journal of Fourier Analysis and Applications, 14 (2008), pp. 877–905.

[21] S. CERIA, *Robust portfolio construction*, 2006. Presentation at Workshop on Mixed Integer Programming, University of Miami, June 5-8, 2006.

[22] S. CERIA AND R. A. STUBBS, *Incorporating estimation errors into portfolio selection: Robust portfolio construction*, Journal of Asset Management, 7 (2006), pp. 109–127.

[23] V. CHANKONG AND Y. Y. HAIMES, eds., *Multiobjective Decision Making: Theory and Methodology*, Elsevier Science Publishing Co., New York, NY, USA, 1983.

[24] F. CHARPIN AND D. LACAZE, *Using binary variables to obtain small optimal portfolios*, Journal of Portfolio Management, 34 (2007), pp. 68–72.

[25] B. Choi and J. O. Deasy, *The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning*, Physics in Medicine and Biology, 47 (2002), pp. 3579–3589.

[26] T.-C. Chou, *Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies*, Pharmacological Reviews, 58 (2006), pp. 621–681.

[27] M. Chu, Y. Zinchenko, S. G. Henderson, and M. B. Sharpe, *Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty*, Physics in Medicine and Biology, 50 (2005), pp. 5463–5477.

[28] T. F. Coleman, J. Henninger, and Y. Li, *Minimizing tracking error while restricting the number of assets*, Journal of Risk, 8 (2006), pp. 33–56.

[29] M. Colombo, *Advances in Interior Point Methods for Large-Scale Linear Programming*, Ph.D. Thesis, School of Mathematics, The University of Edinburgh, Edinburgh, UK, 2007.

[30] G. Cornuejols and R. Tütüncü, *Optimization Methods in Finance*, Mathematics, Finance and Risk, Cambridge University Press, Cambridge, UK, 2006.

[31] D. L. Craft, T. F. Halabi, and T. R. Bortfeld, *Exploration of tradeoffs in intensity-modulated radiotherapy*, Physics in Medicine and Biology, 50 (2005), pp. 5857–5868.

[32] D. L. Craft, T. F. Halabi, H. A. Shih, and T. R. Bortfeld, *Approximating convex Pareto surfaces in multiobjective radiotherapy planning*, Medical Physics, 33 (2006), pp. 3399–3407.

[33] Credit Suisse Financial Products, *CreditRisk+: A credit risk management framework*, Technical Report, Credit Suisse First Boston International, London, UK, 1997.

[34] G. B. DANTZIG, *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ, USA, 11^{th} ed., 1998.

[35] S. DAUL AND E. G. VIDAL, *Replication of insurance liabilities*, RiskMetrics Journal, 9 (2009), pp. 79–96.

[36] M. DE BERG, M. VAN KREVELD, M. OVERMARS, AND O. SCHWARZKOPF, *Computational Geometry: Algorithms and Applications*, Springer-Verlag, Berlin, Germany, 2^{nd} ed., 2000.

[37] A. DE SERVIGNY AND O. RENAULT, *Measuring and Managing Credit Risk*, McGraw-Hill, New York, NY, USA, 2004.

[38] J. O. DEASY, A. I. BLANCO, AND V. H. CLARK, *CERR: A computational environment for radiotherapy research*, Medical Physics, 30 (2003), pp. 979–985.

[39] R. DEMBO AND D. ROSEN, *The practice of portfolio replication. A practical overview of forward and inverse problems*, Annals of Operations Research, 85 (1999), pp. 267–284.

[40] W. S. DORN, *Duality in quadratic programming*, Quarterly of Applied Mathematics, 18 (1960), pp. 155–162.

[41] M. EHRGOTT, *Multicriteria Optimization*, Springer, Berlin, Germany, 2^{nd} ed., 2005.

[42] ——, *Multiobjective optimization*, AI Magazine, 29 (2008), pp. 47–57.

[43] M. EHRGOTT, Ç. GÜLER, H. W. HAMACHER, AND L. SHAO, *Mathematical optimization in intensity modulated radiation therapy*, A Quarterly Journal of Operations Research, 6 (2008), pp. 199–262.

[44] M. EHRGOTT AND M. M. WIECEK, *Mutiobjective programming*, in Multiple Criteria Decision Analysis: State of the Art Surveys, J. Figueira,

S. Greco, and M. Ehrgott, eds., vol. 78 of International Series In Operations Research & Management Science, Springer, New York, NY, USA, 2005, chapter 17, pp. 667–708.

[45] A. S. El-Bakry, R. A. Tapia, and Y. Zhang, *A study of indicators for identifying zero variables in interior-point methods*, SIAM Review, 36 (1994), pp. 45–72.

[46] B. Emami, J. Lyman, A. Brown, L. Coia, M. Goitein, J. E. Munzenrider, B. Shank, L. J. Solin, and M. Wesson, *Tolerance of normal tissue to therapeutic irradiation*, International Journal of Radiation Oncology, Biology, Physics, 21 (1991), pp. 109–122.

[47] F. J. Fabozzi, P.N. Kolm, D.A. Pachamanova, and S.M. Focardi, *Robust Portfolio Optimization and Management*, John Wiley & Sons, Hoboken, NJ, USA, 2007.

[48] J. J. Faraway, *Linear Models with R*, Chapman & Hall/CRC, Boca Raton, FL, USA, 2004.

[49] D. E. Farrar, *The Investment Decision Under Uncertainty*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1962.

[50] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*, IEEE Journal of Selected Topics in Signal Processing, 1 (2007), pp. 586–597.

[51] J. Fliege, *An efficient interior-point method for convex multicriteria optimization problems*, Mathematics of Operations Research, 31 (2006), pp. 825–845.

[52] J. Fliege and A. Heseler, *Constructing approximations to the efficient set of convex quadratic multiobjective problems*, Ergebnisberichte Ange-

wandte Mathematik 211, Fachbereich Mathematik, Universitat Dortmund, Dortmund, Germany, 2002.

[53] E. I. GEORGE, *The variable selection problem*, Journal of the American Statistical Association, 95 (2000), pp. 1304–1308.

[54] A. GHAFFARI-HADIGHEH, H. GHAFFARI-HADIGHEH, AND T. TERLAKY, *Bi-parametric optimal partition invariancy sensitivity analysis in linear optimization*, Central European Journal of Operations Research, 16 (2008), pp. 215–238.

[55] A. GHAFFARI-HADIGHEH, O. ROMANKO, AND T. TERLAKY, *Sensitivity analysis in convex quadratic optimization: Simultaneous perturbation of the objective and right-hand-side vectors*, Algorithmic Operations Research, 2 (2007), pp. 94–111.

[56] ——, *On bi-parametric programming in quadratic optimization*, in Proceedings of the 20th International Conference "Continuous Optimization and Knowledge-Based Technologies", L. Sakalauskas, G. Weber, and E. K. Zavadskas, eds., Neringa, Lithuania, May 20-23, 2008, Vilnius Gediminas Technical University, pp. 229–234.

[57] ——, *Bi-parametric convex quadratic optimization*, Optimization Methods and Software, 25 (2010), pp. 229–245.

[58] ——, *Multiobjective optimization via parametric optimization: Models, algorithms and applications*, in Handbook of Multicriteria Analysis, C. Zopounidis and P. Pardalos, eds., 2010. Forthcoming.

[59] C. J. GOH AND X. Q. YANG, *Analytic efficient solution set for multicriteria quadratic programs*, European Journal of Operational Research, 92 (1996), pp. 166–181.

[60] D. GOLDFARB AND K. SCHEINBERG, *On parametric semidefinite programming*, Applied Numerical Mathematics, 29 (1999), pp. 361–377.

[61] J. GONDZIO AND A. GROTHEY, *A new unblocking technique to warm-start interior point methods based on sensitivity analysis*, SIAM Journal on Optimization, 19 (2008), pp. 1184–1210.

[62] J. GOTOH AND A. TAKEDA, *On the role of norm constraints in portfolio selection*, Discussion Paper Series ISE 09-03, Department of Industrial and Systems Engineering, Chuo University, Tokyo, Japan, 2009.

[63] O. GRODZEVICH AND O. ROMANKO, *Normalization and other topics in multi-objective optimization*, in Proceedings of the First Fields-MITACS Industrial Problems Workshop, D. A. Aruliah and G. M. Lewis, eds., Toronto, Canada, August 2006, Fields Institute for Research in Mathematical Sciences, pp. 89–101.

[64] J. GUDDAT, *Stability in convex quadratic programming*, Mathematische Operationsforschung und Statistik, 7 (1976), pp. 223–245.

[65] J. GUDDAT, F. G. VASQUEZ, K. TAMMER, AND K. WENDLER, *Multiobjective and stochastic optimization based on parametric optimization*, Mathematical Research, 26, Akademie-Verlag, Berlin, Germany, 1985.

[66] O. GÜLER AND Y. YE, *Convergence behavior of interior-point algorithms*, Mathematical Programming, 60 (1993), pp. 215–228.

[67] G. M. GUPTON, C. C. FINGER, AND M. BHATIA, *CreditMetrics – technical document*, Technical Report, RiskMetrics Group, J.P.Morgan, New York, NY, USA, 1997.

[68] T. HESTERBERG, N. H. CHOI, L. MEIER, AND C. FRALEY, *Least angle and ℓ_1 penalized regression: A review*, Statistics Surveys, 2 (2008), pp. 61–93.

[69] M. HLADÍK, *Multiparametric linear programming: Support set and optimal partition invariancy*, European Journal of Operational Research, 202 (2010), pp. 25–31.

[70] A. L. HOFFMANN, D. DEN HERTOG, A. Y. SIEM, J. H. KAANDERS, AND H. HUIZENGA, *Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects*, Physics in Medicine and Biology, 53 (2008), pp. 6345–6362.

[71] A. L. HOFFMANN, A. Y. SIEM, D. DEN HERTOG, J. H. KAANDERS, AND H. HUIZENGA, *Derivative-free generation and interpolation of convex Pareto optimal IMRT plans*, Physics in Medicine and Biology, 51 (2006), pp. 6349–6369.

[72] H. HOLLATZ AND H. WEINERT, *Ein Algorithums zur Lösung des doppelt-einparametrischen linearen Optimierungsproblems*, Mathematische Operationsforschung und Statistik, 2 (1971), pp. 181–197.

[73] T. ILLÉS, J. PENG, C. ROOS, AND T. TERLAKY, *A strongly polynomial rounding procedure yielding a maximally complementary solution for $P_*(\kappa)$ linear complementarity problems*, SIAM Journal of Optimization, 11 (2000), pp. 320–340.

[74] ILOG INC., *ILOG CPLEX User's Manual*, 2008. CPLEX 11.2.

[75] I. ISCOE AND A. KREININ, *Default boundary problem*, Technical Report, Algorithmics Inc., Toronto, Canada, 2001.

[76] I. ISCOE AND A. KREININ, *Large pool approximations for credit loss distributions*, Technical Report ARPS 08-03, Algorithmics Inc., Toronto, Canada, 2008.

[77] I. ISCOE, A. KREININ, H. MAUSSER, AND O. ROMANKO, *Portfolio credit risk optimization*, Technical Report 09-01, Algorithmics Inc., Toronto, Canada, 2009. Submitted to Journal of Credit Risk.

[78] I. ISCOE, A. KREININ, AND D. ROSEN, *An integrated market and credit risk portfolio model*, Algo Research Quarterly, 2 (1999), pp. 21–37.

[79] B. JANSEN, J. J. DE JONG, C. ROOS, AND T. TERLAKY, *Sensitivity analysis in linear programming: just be careful!*, European Journal of Operational Research, 101 (1997), pp. 15–28.

[80] R. JANSEN AND R. VAN DIJK, *Optimal benchmark tracking with small portfolios*, Journal of Portfolio Management, 28 (2002), pp. 33–39.

[81] E. JOHN AND E. A. YILDIRIM, *Implementation of warm-start strategies in interior-point methods for linear programming in fixed dimension*, Computational Optimization and Applications, 41 (2008), pp. 151–183.

[82] N. KARMARKAR, *A new polynomial-time algorithm for for linear programming*, Combinatorica, 4 (1984), pp. 373–395.

[83] P. KROKHMAL, J. PALMQUIST, AND S. URYASEV, *Portfolio optimization with Conditional Value-at-Risk objective and constraints*, Journal of Risk, 4 (2002), pp. 11–27.

[84] K.-H. KÜFER, M. MONZ, A. SCHERRER, P. SÜSS, F. ALONSO, A. S. A. SULTAN, T. BORTFELD, AND C. THIEKE, *Multicriteria optimization in intensity modulated radiotherapy planning*, in Handbook of Optimization in Medicine, P. M. Pardalos and H. E. Romeijn, eds., Springer, Berlin, Germany, 2009, chapter 5, pp. 123–167.

[85] M. KVASNICA, *Real-Time Model Predictive Control via Multi-Parametric Programming: Theory and Tools*, VDM Verlag, Saarbrücken, Germany, 2009.

[86] M. KVASNICA, P. GRIEDER, M. BAOTIĆ, AND M. MORARI, *Multi-Parametric Toolbox (MPT)*, in Hybrid Systems: Computation and Control, R. Alur and G. J. Pappas, eds., vol. 2993 of Lecture Notes in Computer Science, Springer, Berling, Germany, 2004, pp. 448–462.

[87] N. LARSEN, H. MAUSSER, AND S. URYASEV, *Algorithms for optimization of Value-at-Risk*, in Financial Engineering, E-Commerce and Supply

Chain, P. M. Pardalos and V. K. Tsitsiringos, eds., vol. 70 of Applied Optimization, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002, chapter 8, pp. 129–157.

[88] E. K. Lee and J. O. Deasy, *Optimization in intensity-modulated radiation therapy*, SIAG/OPT Views-and-News, Optimization in Medicine, 17 (2006), pp. 20–32.

[89] J. G. Lin, *Three methods for determining Pareto-optimal solutions of multiple-objective problems*, in Directions in Large-Scale Systems, Y. C. Ho and S. K. Mitter, eds., Plenum Press, New York, NY, USA, 1975, pp. 117–138.

[90] ——, *Proper inequality constraints and maximization of index vectors*, Journal of Optimization Theory and Applications, 21 (1977), pp. 505–521.

[91] M. S. Lobo, L. Vandenberghe, S. P. Boyd, and H. Lebret, *Applications of second-order cone programming*, Linear Algebra and Its Applications, 284 (1998), pp. 193–228.

[92] J. Löfberg, *YALMIP: A toolbox for modeling and optimization in MATLAB*, in Proceedings of the CACSD Conference, Taipei, Taiwan, 2004.

[93] C. L. Mallows, *Some comments on Cp*, Technometrics, 15 (1973), pp. 661–675.

[94] R. Mansini, W. Ogryczak, and M. G. Speranza, *LP solvable models for portfolio optimization: a classification and computational comparison*, IMA Journal of Management Mathematics, 14 (2003), pp. 187–220.

[95] H. M. Markowitz, *Portfolio selection*, The Journal of Finance, 7 (1952), pp. 77–91.

[96] ——, *The optimization of a quadratic function subject to linear constraints*, Naval Research Logistics Quarterly, 3 (1956), pp. 111–133.

[97] H. MAUSSER AND D. ROSEN, *Efficient risk/return frontiers for credit risk*, Journal of Risk Finance, 2 (2000), pp. 66–78.

[98] ——, *Applying scenario optimization to portfolio credit risk*, Journal of Risk Finance, 2 (2001), pp. 36–48.

[99] ——, *Scenario-based risk management tools*, in Applications of Stochastic Programming, S. W. Wallace and W. T. Ziemba, eds., MPS-SIAM Series in Optimization, SIAM, Philadeplphia, PA, USA, 2005, chapter 27, pp. 545–574.

[100] L. MCLINDEN, *An analogue of Moreau's proximation theorem, with application to the nonlinear complementary problem*, Pacific Journal of Mathematics, 88 (1980), pp. 101–161.

[101] A. MEUCCI, *Risk and Asset Allocation*, Springer Finance, Springer-Verlag, Berlin, Germany, 1^{st} ed., 2005.

[102] K. M. MIETTINEN, *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers, Boston, MA, USA, 1999.

[103] V. D. MIGUEL, L. GARLAPPI, F. J. NOGALES, AND R. UPPAL, *A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms*, Management Science, 55 (2009), pp. 798–812.

[104] A. MILLER, *Subset Selection in Regression*, Chapman & Hall/CRC, Boca Raton, FL, USA, 2^{nd} ed., 2002.

[105] MOSEK APS, *The MOSEK Optimization Tools Manual. Version 5.0*, 2008. Revision 105.

[106] K. G. MURTY, *Linear Programming*, John Wiley & Sons, New York, NY, USA, 1983.

[107] Y. NESTEROV AND A. NEMIROVSKII, *Interior-Point Polynomial Algorithms in Convex Programming*, vol. 13 of Studies in Applied Mathematics, SIAM, Philadephia, PA, USA, 1994.

[108] A. NIEMIERKO, *A generalized concept of equivalent uniform dose (EUD)*, Medical Physics, 26 (1999), p. 1100.

[109] ——, *Biological optimization*, in Image-Guided IMRT, T. Bortfeld, R. Schmidt-Ullrich, W. D. Neve, and D. E. Wazer, eds., Springer, Berlin, Germany, 2006, chapter II.5, pp. 199–216.

[110] F. NOŽIČKA, *Über eine Klasse vor linearen einparametrischen Optimierungsproblemen*, Mathematische Operationsforschung und Statistik, 3 (1972), pp. 159–194.

[111] F. NOŽIČKA, J. GUDDAT, H. HOLLATZ, AND B. BANK, *Theorie der Linearen Parametrischen Optimierung*, Akademie-Verlag, Berlin, Germany, 1974.

[112] K. NYHOLM, *Strategic Asset Allocation in Fixed-Income Markets: A MATLAB-Based User's Guide*, Wiley Finance Series, John Wiley & Sons, Chichester, UK, 2008.

[113] J. OGRODZKI, *Replicating portfolios*, Insurance and Financial Services Review, Watson Wyatt Limited, Reigate, UK, November 2007.

[114] J. O'ROURKE, *Computational Geometry in C*, Cambridge University Press, Cambridge, UK, 2^{nd} ed., 2001.

[115] A. PAGÈS, J. GONDZIO, AND N. NABONA, *Warmstarting for interior point methods applied to the long-term power planning problem*, European Journal of Operational Research, 197 (2009), pp. 112–125.

[116] E. N. PISTIKOPOULOS, M. C. GEORGIADIS, AND V. DUA, eds., *Multi-Parametric Programming: Theory, Algorithms, and Applications*, vol. 1, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2007.

[117] J.-L. PRIGENT, *Portfolio Optimization and Performance Analysis*, Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall, Boca Raton, FL, USA, 2007.

[118] M. RABINOWITZ, M. BANJEVIC, A. S. CHAN, L. MYERS, R. WOLKOWICZ, J. HABERER, AND J. SINGER, *Use of the l^1 norm for selection of sparse parameter sets that accurately predict drug response phenotype from viral genetic sequences*, in AMIA Annual Symposium Proceedings 2005, 2005, pp. 505–509.

[119] M. RASMUSSEN, *Quantitative Portfolio Optimisation, Asset Allocation and Risk Management*, Palgrave Macmillan, Basingstoke, UK, 2003.

[120] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, USA, 1970.

[121] R. T. ROCKAFELLAR AND S. URYASEV, *Optimization of Conditional Value-at-Risk*, Journal of Risk, 2 (2000), pp. 21–41.

[122] ——, *Conditional Value-at-Risk for general loss distributions*, Journal of Banking and Finance, 26 (2002), pp. 1443–1471.

[123] O. ROMANKO, *An interior point approach to quadratic and parametric quadratic optimization*, M.Sc. Thesis, Department of Computing and Software, McMaster University, Hamilton, Canada, August 2004.

[124] H. E. ROMEIJN AND J. F. DEMPSEY, *Intensity modulated radiation therapy treatment plan optimization*, TOP, 16 (2008), pp. 215–243.

[125] H. E. ROMEIJN, J. F. DEMPSEY, AND J. G. LI, *A unifying framework for multi-criteria fluence map optimization models*, Physics in Medicine and Biology, 49 (2004), pp. 1991–2013.

[126] C. ROOS, T. TERLAKY, AND J.-P. VIAL, *Interior Point Methods for Linear Optimization*, Springer Science, New York, NY, USA, 2006.

[127] RUTTER ASSOCIATES LLC, *Convergence of credit capital models*, Technical Report, Rutter Associates LLC, 2006. Sponsored by the International Association of Credit Portfolio Managers (IACPM) and International Swaps and Derivatives Association (ISDA). http://www.isda.org/c_and_a/pdf/IACPM-ISDA-ConvergenceCreditCapModelsFeb2006.pdf.

[128] S. SATCHELL AND A. SCOWCROFT, eds., *Advances in Portfolio Construction and Implementation*, Butterworth-Heinemann, Burlington, MA, USA, 2003.

[129] A. SAUNDERS AND L. ALLEN, *Credit Risk Mesurement: New Approaches to Value at Risk and Other Paradigms*, John Wiley & Sons, New York, NY, USA, 2^{nd} ed., 2002.

[130] D. SAUNDERS, C. XIOUROS, AND S. A. ZENIOS, *Credit risk optimization using factor models*, Annals of Operations Research, 152 (2007), pp. 49–77.

[131] L. SHAO AND M. EHRGOTT, *Approximately solving multiobjective linear programmes in objective space and an application in radiotherapy treatment planning*, Mathematical Methods of Operations Research, 68 (2008), pp. 257–276.

[132] W. F. SHARPE, *Capital asset prices: a theory of market equilibrium under conditions of risk*, Journal of Finance, 19 (1964), pp. 425–442.

[133] W. F. SHARPE, P. CHEN, J. E. PINTO, AND D. W. MCLEAVEY, *Asset allocation*, in Managing Investment Portfolios: A Dynamic Process, J. L. Maginn, D. L. Tuttle, D. W. McLeavey, and J. E. Pinto, eds., John Wiley & Sons, Hoboken, NJ, USA, 3^{rd} ed., 2007, chapter 5, pp. 230–327.

[134] A. Y. D. SIEM, D. DEN HERTOG, AND A. L. HOFFMANN, *The effect of transformations on the approximation of univariate (convex) functions with applications to Pareto curves*, European Journal of Operational Research, 189 (2008), pp. 347–362.

[135] M. Y. Sir, M. A. Epelman, and S. M. Pollock, *Stochastic programming for off-line adaptive radiotherapy*, Preprint, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA, 2009.

[136] R. E. Steuer, *Multiple Criteria Optimization: Theory, Computation, and Application*, John Wiley & Sons, New York, NY, USA, 1986.

[137] F. J. F. Svetlozar T. Rachev, Stoyan V. Stoyanov, *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization*, John Wiley & Sons, Chichester, UK, 2008.

[138] G. Szegö, ed., *Risk Measures for the 21st Century*, Wiley Finance Series, John Wiley & Sons, Chichester, UK, 2004.

[139] C. Thieke, T. R. Bortfeld, and K.-H. Küfer, *Characterization of dose distribution through the max and mean dose concept*, Acta Oncologica, 41 (2002), pp. 158–161.

[140] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society, Ser. B, 58 (1996), pp. 267–288.

[141] S. Tilke, *Reducing asset weights' volatility by importance sampling in stochastic credit portfolio optimization*, Technical Report 417, Department of Economics, University of Regensburg, Regensburg, Germany, 2006.

[142] P. Tøndel, T. A. Johansen, and A. Bemporad, *An algorithm for multi-parametric quadratic programming and explicit MPC solutions*, Automatica, 39 (2003), pp. 489–497.

[143] A. Wächter and L. T. Biegler, *On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming*, Mathematical Programming, 106 (2006), pp. 25–57.

[144] S. W. WALLACE, *Decision making under uncertainty: is sensitivity analysis of any use*, Operations Research, 48 (2000), pp. 20–25.

[145] H. WANG, G. LI, AND G. JIANG, *Robust regression shrinkage and consistent variable selection via the LAD-Lasso*, Journal of Business & Economics Statistics, 25 (2007), pp. 347–355.

[146] Y. WANG, X. WANG, AND Y. CHENG, *A causal discovery approach to identifying active components of herbal medicine*, in Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, September 2005.

[147] ——, *Empirical study on modeling quantitative composition–activity relationships in chinese herbal medicine*, in Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, September 2005.

[148] ——, *A computational approach to botanical drug design by modeling quantitative composition–activity relationship*, Chemical Biology & Drug Design, 68 (2006), pp. 166–172.

[149] H. WEINERT, *Doppelt-einparametrische lineare Optimierung. I: Unabhängige Parameter*, Mathematische Optimierungsproblems und Statistik, 1 (1970), pp. 173–197.

[150] S. E. WRIGHT, *Identifying the optimal partition in convex quadratic programming*, Operations Research Letters, 36 (2008), pp. 67–70.

[151] Y. YE, *Interior Point Algorithms: Theory and Analysis*, John Wiley & Sons, New York, NY, USA, 1997.

[152] E. A. YILDIRIM, *An interior-point perspective on sensitivity analysis in linear programming and semidefinite programming*, Ph.D. Thesis, School of Operations Research and Information Engineering, Cornell University, Ithaca, NY, USA, 2001.

[153] ——, *Unifying optimal partition approach to sensitivity analysis in conic optimization*, Journal of Optimization Theory and Applications, 122 (2004), pp. 405–423.

[154] K. Yu, Y. Gong, Z. Lin, and Y. Cheng, *Quantitative analysis and chromatographic fingerprinting for the quality evaluation of scutellaria baicalensis georgi using capillary electrophoresis*, Journal of Pharmaceutical and Biomedical Analysis, 43 (2007), pp. 540–548.

[155] L. Yu, X. Ji, and S. Wang, *Stochastic programming models in financial optimization: A survey*, Advanced Modeling and Optimization, 5 (2003), pp. 1–26.

[156] R. Zagst, J. Kehrbaum, and B. Schmid, *Portfolio optimization under credit risk*, Computational Statistics, 18 (2003), pp. 317–338.

[157] S. A. Zenios, ed., *Financial Optimization*, Cambridge University Press, Cambridge, UK, 1996.

[158] S. A. Zenios and N. J. Jobst, *The tail that wags the dog: Integrating credit risk in asset portfolios*, Journal of Risk Finance, 3 (2001), pp. 31–43.

[159] Y. Zinchenko, T. Craig, H. Keller, T. Terlaky, and M. B. Sharpe, *Controlling the dose distribution with gEUD-type constraints within the convex radiotherapy optimization framework*, Physics in Medicine and Biology, 53 (2008), pp. 3231–3250.

[160] C. Zopounidis, D. K. Despotis, and I. Kamaratou, *Portfolio selection using the ADELAIS multiobjective linear programming system*, Computational Economics, 11 (1998), pp. 189–204.

Index