

SOLVING A QUADRATIC PROGRAMMING PROBLEM SUBJECT TO ORTHOGONALITY CONSTRAINTS

By

Stephen E. Braun

A Thesis Submitted to the Graduate
Faculty of Rensselaer Polytechnic Institute
in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
Major Subject: Mathematical Sciences

Approved by the
Examining Committee:

Professor John E. Mitchell, Thesis Adviser

Kristin P. Bennett, Member

Joseph G. Ecker, Member

Albert S. Paulson, Member

Rensselaer Polytechnic Institute
Troy, New York

November 11, 2001
(For Graduation December 2001)

SOLVING A QUADRATIC PROGRAMMING PROBLEM SUBJECT TO ORTHOGONALITY CONSTRAINTS

By

Stephen E. Braun

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Mathematical Sciences

The original of the complete thesis is on file
in the Rensselaer Polytechnic Institute Library

Examining Committee:

Professor John E. Mitchell, Thesis Adviser

Kristin P. Bennett, Member

Joseph G. Ecker, Member

Albert S. Paulson, Member

Rensselaer Polytechnic Institute
Troy, New York

November 11, 2001
(For Graduation December 2001)

© Copyright 2001
by
Stephen E. Braun
All Rights Reserved

CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
ACKNOWLEDGMENT	ix
ABSTRACT	x
1. Introduction	1
1.1 Background and Motivation	1
1.2 Standard Form Quadratic Program	2
1.3 Orthogonality Constraints on Continuous Variables	3
1.3.1 Selected Modeling Examples: (SOS1), (SOS2), and (DOP)	3
1.4 Classification of (QPO) and Ideas for Solution	6
1.4.1 A Brief Introduction to Semidefinite Programming	6
1.4.2 Usefulness of Semidefinite Programming	7
1.4.3 Capabilities of Available SDP Solvers	9
1.5 Organization of Remaining Chapters	10
2. The Restatement of (QPO)	13
2.1 Step 1: Shift to Zero Lower Bound	14
2.2 Step 2: Introduce Slacks into Inequalities	16
2.3 Step 3: Incorporate Linear Term in Quadratic Objective	17
2.4 Step 4: Homogenize Equalities	18
3. The Transition from (QPO) to (rSDP)	20
3.1 Further Remarks on Semidefinite Programming	20
3.2 Implications of $x, y \longrightarrow Z$	21
3.3 Imposing Orthogonality Constraints	22
3.4 Imposing Bound Constraints	22
3.5 Imposing $y = 1$	24
3.6 Symmetric Lifting of Homogenized Linear Equalities	25
3.6.1 Lifting Example	26
3.6.2 Alternate Analytic Construction for $Az = 0$	28

3.7	Relaxing (QPO) into (rSDP)	29
3.8	Relating the Sizes of (QPO) and (rSDP)	29
4.	(DOP) Solution Strategy and Test Problems	31
4.1	Special Structure of (DOP) Orthogonality	31
4.2	Remarks on Terminology	31
4.3	Simple Example: (QPO) with (DOP) Orthogonality	32
4.3.1	The Naïve Approach for Solution	34
4.3.2	Usefulness of Accurate Active Set Information	35
4.4	Our Solution Strategy: (rSDP) + QP-subproblems	35
4.5	Description as an Enumerative Algorithm	37
4.6	Deduction of Optimal Active Set Information	39
4.6.1	Estimating x^* from Z^*	39
4.6.2	Decision Criteria for Active Set Information	41
4.7	Creation of Test Problems and Results	44
4.7.1	Linear Programming Problems	44
4.7.2	Random Symmetric Quadratic Test Problem	45
5.	Application of (QPO): The Portfolio Rebalancing Problem	48
5.1	The Standard Portfolio Problem and Extensions	48
5.2	Inclusion of Transaction Costs	49
5.2.1	Motivation for Percentage Transaction Cost Model	50
5.2.2	Rebalancing a Portfolio under Percentage Transaction Costs	51
5.3	Effect of Changing Parameters	53
5.4	Crossover Among “Efficient” Frontiers	54
5.4.1	Necessity of (DOP) Orthogonality	54
5.4.2	Nonconvexity of the Feasible Region	57
5.4.3	Issue of the Remaining Crossover: Fractional Objective	57
5.5	Performance Analysis for the Markowitz $c=5\%$ TCEF	60
5.5.1	Examination of the QP-subproblems Selected by the Strategy	62
5.5.2	Presentation of the Markowitz $c=5\%$ TCEF Portfolios	65
5.6	Performance Analysis for an Additional Markowitz TCEF	68
5.7	Rebalancing a Dow Jones Portfolio	72

6. The Issue of Degenerate Decisions	79
6.1 Respond by Growing the Search Space	79
6.2 Visualizing the Geometry	80
6.3 Identifying the Neighboring Faces	85
6.4 Generality of the Tree Representation	87
6.5 Ways to Grow the Search Space	89
6.5.1 The Conservative Choice	91
6.5.2 A More Aggressive Approach	91
6.6 Effect of NP-hardness on the Strategy	92
7. Final Discussion and Concluding Remarks	95
7.1 Summary	95
7.2 Main Contributions	95
7.2.1 Finding a New Role for y	96
7.2.2 The Symmetric Lifting Procedure	96
7.2.3 Scaling the Risk Measurement	97
7.3 Directions for Future Work	97
7.3.1 Multi-period Rebalancing	97
7.3.2 Explicitly Constraining the Costs Paid to Rebalance	98
7.3.3 Pursuing Additional Applications	98
7.3.4 Improving the Implementation	99
LITERATURE CITED	100
APPENDICES	104
A. Portfolio Datasets Used	104
B. Bounding the Portfolio Expected Return	106
C. Transforming the Fractional Objective	108

LIST OF TABLES

3.1	Relating the Dimensionalities of (QPO) and (rSDP)	30
5.1	Performance of the Strategy for the Markowitz $c=5\%$ TCEF	62
5.2	Distribution of QP-subproblems Selected by the Strategy	63
5.3	Comparison of Active Set Choices for QP-subproblems	65
5.4	Parameters used for the Additional Markowitz TCEF	68
5.5	Performance of the Strategy for the Additional Markowitz TCEF	70
5.6	Parameters used for the Dow Jones Portfolio	73
5.7	Performance of the Strategy for the Dow Jones Portfolio	75
A.1	The Markowitz Dataset	104
A.2	The Dow Jones Dataset	105

LIST OF FIGURES

2.1	Graph Structure of Orthogonality Constraints	14
4.1	Solution Strategy Flowchart	36
4.2	Tree Representation of the Enumerative Algorithm	38
4.3	Illustration of Percent Upper Bound Criteria.	42
4.4	Test Problem Performance of the Strategy	47
5.1	Multiple “Efficient” Frontiers using Same Initial Portfolio	54
5.2	Structure of Orthogonality in the Portfolio Rebalancing Problem	56
5.3	Nonconvexity of the Feasible Region in Two Dimensions	58
5.4	Multiple Efficient Frontiers using Same Initial Portfolio	59
5.5	Computational Results for the Markowitz $c=5\%$ TCEF	61
5.6	Comparison of QP-subproblem Optimal Solutions	64
5.7	Visualization of the Involved Securities for the Markowitz $c=5\%$ TCEF	66
5.8	Computational Results for the Additional Markowitz TCEF	69
5.9	Involved Securities for the Additional Markowitz TCEF	71
5.10	Computational Results for the Dow Jones Portfolio	74
5.11	Involved Securities for the Dow Jones Portfolio: $c=0\%$ Frontier	76
5.12	Involved Securities for the Dow Jones Portfolio: TCEF	77
6.1	Convex Feasible Region in Three Dimensions	81
6.2	Projections of the Convex Feasible Region	82
6.3	Nonconvex Feasible Region in Three Dimensions	83
6.4	Projections of the Nonconvex Feasible Region	84
6.5	Geometric Relationship between Degeneracy and Neighboring Faces	86
6.6	Illustration of the Effect of Degeneracy on the Tree	88
6.7	Transitioning to the Tree Representation	90

6.8	Example of Aggressive Approach: Initial Stage	93
6.9	Example of Aggressive Approach: Response to Degeneracy	94

ACKNOWLEDGMENT

Thanks to

my family,

John Mitchell,

my committee,

Susan Cole,

friends, faculty and staff of the Mathematical Sciences Department,

and the many teachers I have learned from.

ABSTRACT

This thesis considers quadratic programming problems where combinatorial constraints are directly imposed on the continuous decision variables using a set of pairwise orthogonality relationships; we denote this problem (QPO). These orthogonality constraints are non-convex in nature, meaning the resulting combinatorial optimization problem is NP-hard. The addition of orthogonality constraints on continuous variables is shown to encompass a wide range of modeling choices.

Traditional approaches for accommodating such combinatorial constraints have been to introduce additional binary variables and solve the resulting mixed-integer programming problem. Here we instead construct a semidefinite programming problem relaxation for (QPO); we denote this relaxation (rSDP). For (rSDP), a symmetric lifting procedure for homogenized linear equalities has been developed. With a general set of orthogonality relationships, the optimal objective function value of (rSDP) serves as a lower bound on (QPO). Also, placing (rSDP) into an enumerative algorithm is shown to be a useful strategy for limiting the search space.

Finally a financial application of (QPO), the portfolio rebalancing problem in the presence of transaction costs, is fully explored and computational results are presented. In this application, pairwise orthogonality constraints are imposed between buying and selling decisions. Information about the optimal portfolio is deduced from the (rSDP) solution matrix and the geometry of the feasible region.

CHAPTER 1

Introduction

Before progressing into the mathematical content of this thesis, it is important to first provide some context and motivation. This thesis grew out of research in the area of financial optimization. As the name suggests, financial optimization deals with the application or development of mathematical programming techniques for financial decision-making. Evidence of vigorous research activity within this field is easy to document. Recently, an entire issue of *Mathematical Programming* was devoted to this emerging field of study [34]. Among the topics appearing in that single volume were: dynamic asset-liability management [51], index-tracking [13], credit risk management [4], and portfolio selection [26]. This thesis presents a new approach and strategy applicable to an extension of the portfolio selection problem.

1.1 Background and Motivation

One of the most significant financial decisions facing individuals and institutions is the construction of optimal investment portfolios. These investment decisions are made in the present; their ultimate success or failure is only realized in the future. So this type of decision takes place under uncertainty, a common theme for financial optimization. One response to this atmosphere of uncertainty is to base judgments regarding optimality on current expectations of future returns and current perceptions of future risk.

The central question of the portfolio problem is: What is the optimal way to allocate a limited amount of capital amongst a given set of investment choices? To answer this question, a process must be developed that can determine the relative weight each investment choice should have within the portfolio. That allocation should strike an acceptable balance between risk and reward. In addition, costs incurred when setting up a new portfolio or changing an existing portfolio are an important consideration. These costs must be included in any realistic decision-making process.

This thesis presents an application of mathematical programming, specifically semidefinite programming, which provides useful information to aid decision-makers in answering this question. It is the work of subsequent sections and chapters to show exactly how a topic such as semidefinite programming can be productively applied to this type of financial decision.

1.2 Standard Form Quadratic Program

Beginning our mathematical development, quadratic programming problems (QP) are a fundamental type of math programming problem. Many direct applications exist. Also one of the approaches for nonlinear optimization, SQP methods, relies on solving multiple locally approximate QP.

The standard form quadratic programming problem involves a quadratic objective with a linear term. There are a finite number of continuous decision variables. These variables are subject to bound, linear inequality, and linear equality constraints. One canonical way to state this problem is:

$$\begin{aligned}
 \text{(QP)} \quad & \min_x && x^T Q x + c^T x \\
 & \text{s.t.} && A_{ieq} x \leq b_{ieq} \\
 & && A_{eq} x = b_{eq} \\
 & && l \leq x \leq u
 \end{aligned}$$

It is worth spending a moment to focus more explicitly on the vector variable, $x \in \mathbb{R}^n$. Standard QPs involve a finite number of continuous variables; the components of x constitute a set of variables $\{x_1, x_2, \dots, x_n\}$. These variables are individually subject to independent upper and lower bounds. They are linked together by the linear equality and inequality constraints. However there is no explicit combinatorial structure present in the bound or linear constraints. For example, one could not directly impose a constraint such as:

$$\textit{Either } x_1 \textit{ can be nonzero or } x_2 \textit{ can be nonzero but not both.}$$

What we consider is the addition of a specific type of combinatorial constraint to the standard QP. Our constraints have the form $x_i x_j = 0$ for some specified

set of pairs of variables $\{x_i, x_j\}$. We call this type of constraint an orthogonality constraint and a quadratic optimization problem involving these constraints (QPO); for our standard form, see Chapter 2. Continuing this introduction, we turn to the combinatorial nature and modeling flexibility of orthogonality constraints.

1.3 Orthogonality Constraints on Continuous Variables

To illustrate the combinatorial nature of an orthogonality constraint consider the following basic example:

Let $x \in \mathbb{R}_+^2$.

Impose the constraint $x_1 x_2 = 0$.

Notice that this orthogonality constraint allows for three possible cases:

$$\left\{ \begin{array}{l} x_1 > 0 \\ x_2 = 0 \end{array} \right\} \quad \left\{ \begin{array}{l} x_1 = 0 \\ x_2 > 0 \end{array} \right\} \quad \left\{ \begin{array}{l} x_1 = 0 \\ x_2 = 0 \end{array} \right\}$$

Two of these cases are definitive in the sense that exactly one member of the orthogonal pair of variables is nonzero. The third case is degenerate in the sense that both members of the orthogonal pair are zero. Collectively, these three cases imply that either x_1 can be nonzero or x_2 can be nonzero but not both. So there is a clear combinatorial structure to an orthogonality constraint.

With this orthogonality constraint, the set of variables $\{x_1, x_2\}$ can be classified as a special ordered set of type-I, (SOS1). This introduces a set having a general type of combinatorial property; others will be discussed in the following section.

1.3.1 Selected Modeling Examples: (SOS1), (SOS2), and (DOP)

As previously introduced, (SOS1) is a general type of combinatorial constraint that can be imposed on a set of variables. We will show how this type of combinatorial constraint can be accommodated in (QPO). All modeling examples in this section assume $x \in \mathbb{R}_+^n$. A verbal description of (SOS1) is that at most one member

of the set is nonzero. Analytically, this restriction is expressed by the following set of orthogonality constraints: $x_i x_j = 0$ whenever $i \neq j$.

The following instance of (QPO) incorporates (SOS1):

$$\begin{aligned}
 & \min_x x^T Q x + c^T x \\
 & \text{s.t. } A_{ieq} x \leq b_{ieq} \\
 & \quad A_{eq} x = b_{eq} \\
 & \quad l \leq x \leq u \\
 & \quad \left\{ \begin{array}{l} x_i x_j = 0 \quad \forall i \neq j \\ \text{for } n_1 \leq i, j \leq n_2 \end{array} \right\}
 \end{aligned}$$

Allowing $n_1 \leq n_2 \leq n$ only hints at the range of modeling opportunities. In fact, multiple “type-I” subsets could be specified. Further, orthogonally unconstrained variables can be present. General linear constraints involving all variables or perhaps just involving one or more of the type-I subsets can be supplied with generality. This notation is admittedly clumsy; however, a better representation for a general set of orthogonality constraints will soon be presented.

An existing technique for handling a (SOS1) combinatorial constraint is to introduce additional binary variables, impose additional constraints, and then solve the resulting mixed-integer programming problem (MIP), see [35].

For example, let $\{x_1, x_2, \dots, x_n\}$ be a (SOS1), where $0 \leq x_i \leq u_i$.

To enforce this, introduce n new binary variables $\{y_1, y_2, \dots, y_n\}$.

The necessary constraints are: $x_i \leq u_i y_i \forall i$ and $\sum_{i=1}^n y_i \leq 1$.

This approach imposes the same “type-I” combinatorial constraint on x , respecting its upper bound, by turning the problem into a MIP. Notice that the additional constraints required involve both the continuous variable x and discrete variable y . Recent work by de Farias et al. incorporates knowledge of combinatorial constraints directly into a branch-and-cut approach for solving such a MIP where the objective is linear [14].

Moving now to a second modeling example, consider a special ordered set of type-II, (SOS2), see [16]. The combinatorial constraint here is that at most both members of a single adjacent pair of variables can be nonzero. The following set of orthogonality constraints enforces exactly this condition:

$$x_i x_j = 0 \text{ whenever } |i - j| \geq 2.$$

As with (SOS1), modeling possibilities include one or more “type-II” subsets. Also, both type-I and type-II subsets are allowed as well as the involvement of orthogonally unconstrained variables. This combinatorial structure can be imposed in addition to whatever bound and/or linear constraints may be needed for a specific problem. The idea of special ordered sets can be generalized to greater cardinalities. For our purposes, we simply state that these combinatorial constraints can be accommodated within (QPO).

The final modeling example will be central to the portfolio rebalancing application presented in Chapter 5. In this specialization of the multiple “type-I” subset case, orthogonality relationships exist only between disjoint pairs of variables. We will denote this type of orthogonality by (DOP). A more formal way to state this (DOP) restriction is:

$$\begin{aligned} & \text{If } x_i x_j = 0 \text{ then neither } x_i \text{ nor } x_j \text{ can be} \\ & \text{orthogonally constrained with any other variable.} \end{aligned}$$

Note that for $x \in \mathbb{R}_+^n$, there can be at most $\lfloor \frac{n}{2} \rfloor$ disjoint orthogonal pairs. Unlike a general set of orthogonality relationships, the (DOP) case separates the larger combinatorial problem into consideration of each orthogonal pair of variables. Here when $x_i x_j = 0$ then, for the definitive cases there are only two choices, either $x_i > 0$ or $x_j > 0$, not both. This is an important point which will be further elaborated on in Section 4.1.

These three examples are clearly not exhaustive; other modeling choices are possible. However this small selection was intended to demonstrate the usefulness of adding orthogonality constraints on continuous variables to the repertoire of math programming constraints. It is also important to stress that the orthogonality relationships can be general and have a clear interpretation as combinatorial constraints. These combinatorial constraints work in concert with the usual bound and linear

constraints. Also, the entire problem is stated within the context of minimizing a quadratic objective. As will be presented in subsequent chapters, we can construct a relaxation of (QPO) for a general set of orthogonality constraints.

1.4 Classification of (QPO) and Ideas for Solution

One observation, intentionally not made before now, is that the orthogonality constraints under consideration have an especially simple quadratic structure. This places our problem under the general classification of Q²P, that is quadratically constrained quadratic programming problems. Such problems are known in general to be NP-hard. However, these problems have been successfully approached using semidefinite programming [28, 39].

An aside: (QPO) relates to what are known as complementarity problems and in particular to a type of problem known as a mathematical programming problem with equilibrium constraints, abbreviated MPEC [30]. To illustrate, a common constraint found in MPEC is:

$$0 \leq x \perp F(x) \geq 0$$

where F is a vector-valued function. This can also be stated as:

$$x \geq 0, F(x) \geq 0, x_i F_i(x) = 0, \forall i.$$

Here if F is linear or if a linearization is appropriate, this type of constraint can be incorporated into an instance of (QPO) with (DOP) orthogonality. Therefore the solution technique presented in this thesis may be useful more generally.

1.4.1 A Brief Introduction to Semidefinite Programming

Semidefinite Programming (SDP) is an important and active research area; for a thorough introduction to SDP, see [46] and the references therein. SDP has applications in control theory, eigenvalue optimization, structural design, among others. In addition, SDP relaxations of difficult graph theoretic and combinatorial problems have been formulated using continuous variables to relax discrete $\{-1, 1\}$ or $\{0, 1\}$ variables. These relaxations are solvable in polynomial time. The fact that others had applied SDP to combinatorial problems and obtained polynomial-time solutions initially sparked our interest in this area.

To assist the reader by making this thesis more self-contained, we will summarize SDP notation and survey some basic results. Perhaps because of the variety of disciplines where SDP has found application, each author seems to have their own “standard form” for SDPs. We will not add yet another to the mix. Instead we cite [43] and use the following standard forms.

$$\begin{array}{ll}
 \textbf{Primal SDP} & \min_Z \quad C \bullet Z \\
 & \text{s.t.} \quad A_k \bullet Z = b_k \quad \text{for } k = 1, \dots, K \\
 & \quad \quad Z \succeq 0
 \end{array}$$

Here C and A_k are symmetric matrices, b_k are scalars, and Z is a positive semidefinite matrix variable as denoted by $Z \succeq 0$. For reference, the Frobenius matrix product has the following definition: $C \bullet Z = \text{trace}(CZ) = \sum_{i,j} C_{ij}Z_{ij}$.

$$\begin{array}{ll}
 \textbf{Dual SDP} & \max_{y,S} \quad b^T y \\
 & \text{s.t.} \quad \sum_{k=1}^K y_k A_k + S = C \\
 & \quad \quad S \succeq 0
 \end{array}$$

Here $b \equiv [b_1 \ b_2 \ \dots \ b_K]^T$, S is a slack matrix variable, and y is a vector variable.

The duality theory for semidefinite programming is closely related to that of linear programming. Weak duality is satisfied, namely:

$$C \bullet Z - b^T y = Z \bullet S \geq 0.$$

However, strong duality does not exist in general. In other words for a general SDP, the gap between optimal primal and dual objective values can be nonzero. With the assumption of a Slater (strict feasibility) condition, strong duality does hold and the analogy to linear programming is repaired.

1.4.2 Usefulness of Semidefinite Programming

As was mentioned previously, SDP relaxations of difficult graph theoretic problems have been formulated. The work of Goemans and Williamson with MAXCUT is a classic result of this type, see [21]. They were able to establish an analytic worst-case performance bound of .87856... on their SDP relaxation of MAXCUT. Their

results have since been extended so that weaker worst-case performance bounds now exist for other related problems. For example, the SDP relaxation of an instance of Q²P constrained by m symmetric positive semidefinite quadratic constraints has been shown to be bounded in terms of relative accuracy by $1 - \frac{1}{2 \ln(2m^2)}$. This result and an overview of other extensions of the work of Goemans and Williamson can be found in [36].

In conjunction with their analytic bound, Goemans and Williamson developed a randomized approximation scheme to recover discrete-valued solution from the solution of the continuous SDP relaxation. In (QPO) however, the continuous variables themselves are directly important. Without the need to recover discrete variables, such randomization schemes are not applicable although we will use the solution of (rSDP) to deduce membership in the optimal active set.

Other problems drawn from Q²P with properties closer in appearance to (QPO) have been researched. Properties of their SDP relaxations have been identified and in some instances polynomial-time algorithms have been authored. For example, Zhang considered a class of quadratic maximization problems allowing for sign restrictions of the form:

$$x_i x_j \geq 0 \text{ and } x_k x_l \leq 0.$$

There the author precludes what we call orthogonality constraints by assumption, see [50]. In other words, the implication of equality arising directly from opposing double inequalities is not allowed.

Nesterov has examined the effect of a single linear constraint on the SDP relaxation of a quadratic programming problem [37]. Ye and Zhang [49] have gone further and allow for a set of linear complementarity constraints having the following form:

$$\begin{aligned} \bar{a}_i^T x &\leq a_{i0}, \quad \text{for } i = 1, \dots, m \\ (a_{i0} - \bar{a}_i^T x) (a_{j0} - \bar{a}_j^T x) &= 0 \text{ for all } i \neq j. \end{aligned}$$

However, (QPO) differs from all of these problems in the general type of nonconvexity allowed in the feasible region. This nonconvexity will be highlighted in Chapter 6.

As an aside, Wolkowicz in his discussion of semidefinite relaxations for the MAXCLIQUE problem states orthogonality constraints on binary variables rep-

representing edges in a graph [38]. His association of such constraints with edges in a graph motivates our use of a graph in the standard form of (QPO). This is the improved representation for a general set of orthogonality constraints that was promised in Section 1.3.1. The adjacency matrix of such a graph provides an easy way to reference any set of orthogonality constraints. In addition, (DOP) orthogonality can be thought of as limiting our consideration to only simple graphs.

1.4.3 Capabilities of Available SDP Solvers

Since semidefinite programming problems are useful for a wide range of disciplines and are solvable in polynomial time, it should be no surprise that significant effort has gone into the development of solvers for this type of problem. Some algorithms are tailored to efficiently solve only a specific problem type, for example MAXCUT [10, 47]. Others are able to solve general classes of SDP problems. Of these, several solvers were evaluated [3, 9, 23, 25, 40, 44, 45]. Mittelman has recently compiled a comprehensive survey of the performance of various SDP solvers as part of the Seventh DIMACS Challenge [33].

In this research, two solvers were used extensively. The first was the SDPT3 solver developed by K. C. Toh, M. J. Todd, and R. Tütüncü [44, 45]. This solver makes use of the MATLAB programming environment [32]. In general terms, the SDPT3 solver uses an interior-point approach and iterates in a primal-dual manner, factorizing the Schur complement at each step. This calculation is performed to verify that the solution matrix remains positive semidefinite; however, it is the most time consuming part of each iteration. Background on the Schur complement can be found in texts such as [15, 22]. Commentary on the complexity of interior-point approaches when the number of SDP constraints is large can be found in [27].

The second solver avoids the necessity of factorizing the Schur complement at each iteration by making use of a formulation which relies on properties of the dual SDP. This solver is a C++ implementation of the Spectral Bundle Method made available by C. Helmberg [23, 25]. In briefest terms, the spectral bundle method uses a subgradient approach to solve the nonsmooth convex eigenvalue maximization problem implicitly contained in the dual SDP [24]. At each iteration, the numerically

efficient Lanczos process is used to compute the minimal eigenvalues of the current solution matrix and their corresponding eigenvectors. Negative eigenvalues generate constraints which force positive semidefiniteness on the solution matrix.

The use of multiple solvers allowed for continuous verification of the consistency between results as well as performance comparisons. In terms of performance, though such information becomes almost immediately outdated, the current practical limit on problem size that can be handled by SDPT is on the order of hundreds. This limit applies to both the number of variables that can be handled as well as the number of constraints that can be imposed. Though the spectral bundle method is able to effectively handle problems larger problems, other factors limited our consideration to problems of similar size. The effect of these performance limits on our presentation of computational results is discussed briefly in Section 3.8 and again in Chapter 7.

1.5 Organization of Remaining Chapters

To conclude this introduction, we outline the structure of this thesis. Chapter 2 presents the standard forms that will be used for (QPO) and its semidefinite relaxation, (rSDP). This chapter then progresses through all the preparatory steps that are required before (QPO) can be relaxed into (rSDP). The unique complications arising from orthogonality constraints are treated as they appear.

Chapter 3 details construction of the matrix constraints required for the relaxation of (QPO) into (rSDP). Here, special attention is given to the symmetric lifting procedure developed for the homogenized linear constraints.

Continuing, Chapter 4 returns to the subject of (DOP) orthogonality and describes the solution strategy used in this case. The heuristic deduction of information from the optimal solution of (rSDP) regarding the optimal active set of (QPO) is presented in detail. This chapter concludes with two test problems.

Chapter 5 explores a financial application of (QPO), the portfolio rebalancing problem in the presence of transaction costs. As previously stated, this problem motivated the research presented in this thesis. Our model for transaction costs and the inclusion of transaction costs as an additional “security” with special risk-

return characteristics is explained. The formulation of this problem as an instance of (QPO) is detailed. The remainder of the chapter presents computational results.

Chapter 6 focuses on the issue of degenerate decisions and the importance of understanding the geometry of the feasible region. That geometric insight provides our rationale for growing the search space when degeneracy is identified. To help convey that insight, the feasible region is visualized for the problem of rebalancing a three security portfolio.

Finally Chapter 7 summarizes our work, highlighting the main contributions of this thesis. Also since there is always something new to be learned, directions where continued research can occur are described.

Organizational Note:

Since the applications contained in this thesis are explored numerically and all involve significant computation, it seems to the author that implementation issues deserve at least a brief high-level discussion. Rather than segregate this implementation discussion into a separate chapter, we have chosen to instead highlight the most interesting issues at relevant points along the way.

CHAPTER 2

The Restatement of (QPO)

First, we state the standard form that will be used for a quadratic programming problem with orthogonality constraints (QPO).

$$\begin{aligned}
 \text{(QPO)} \quad & \min_x && x^T Q x + c^T x \\
 & \text{s.t.} && A_{ieq} x \leq b_{ieq} \\
 & && A_{eq} x = b_{eq} \\
 & && l \leq x \leq u \\
 & && x_i x_j = 0 \quad \forall (i, j) \in \Phi
 \end{aligned}$$

Here Q is a real, symmetric matrix and Φ is a graph. Nodes in the graph represent variables and adjacency indicates the presence of an orthogonality constraint on a pair of variables. Figure 2.1 presents an example. The vector s will represent slack variables. The important double role that scalar variable y will play in our development is discussed later in this chapter.

We will be working towards a relaxed semidefinite programming (rSDP) formulation for (QPO). For reference purposes, we include that standard form here as well:

$$\begin{aligned}
 \text{(rSDP)} \quad & \min_Z && C \bullet Z \\
 & \text{s.t.} && A_k \bullet Z = b_k \quad \text{for } k = 1, \dots, K \\
 & && Z \succeq 0
 \end{aligned}$$

Here C and A_k are symmetric matrices and Z is a matrix variable. This choice obviously corresponds to the primal SDP discussed in Chapter 1.

The remaining sections of this chapter present in detail the steps necessary to restate (QPO) in terms suitable for the relaxation. Complications arising from the inclusion of orthogonality constraints are explicitly addressed as they appear. A discussion of the relationship between the dimensionalities of these two problems can be found in Section 3.8.

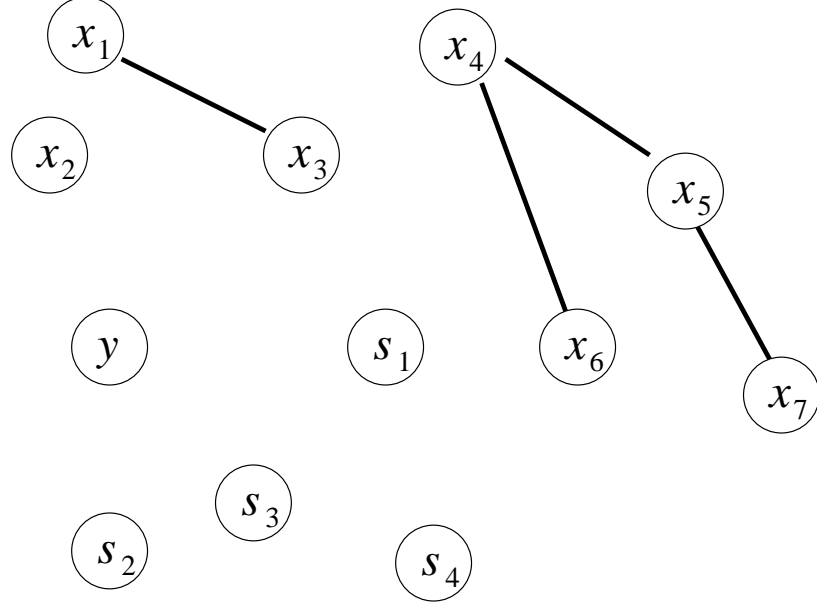


Figure 2.1: This exemplifies the type of graph we are considering. Edges can exist between components of x . However, y and components of s are isolated nodes.

2.1 Step 1: Shift to Zero Lower Bound

Let $\tilde{x} = x - l$ so then $x = \tilde{x} + l$ can be substituted throughout. The first part of the problem to be considered is the objective function.

$$\begin{aligned}
 x^T Q x + c^T x &\longrightarrow (\tilde{x} + l)^T Q (\tilde{x} + l) + c^T (\tilde{x} + l) \\
 &\longrightarrow (\tilde{x} + l)^T Q (\tilde{x} + l) + c^T \tilde{x} + c^T l \\
 &\longrightarrow (\tilde{x}^T + l^T) Q (\tilde{x} + l) + c^T \tilde{x} + c^T l \\
 &\longrightarrow \tilde{x}^T Q \tilde{x} + \tilde{x}^T Q l + l^T Q \tilde{x} + l^T Q l + c^T \tilde{x} + c^T l
 \end{aligned}$$

Using the fact that Q is symmetric and dropping all constant terms yields:

$$\begin{aligned}
 &\longrightarrow \tilde{x}^T Q \tilde{x} + 2l^T Q \tilde{x} + c^T \tilde{x} \\
 &\longrightarrow \tilde{x}^T Q \tilde{x} + (2Ql + c)^T \tilde{x} \\
 &\longrightarrow \tilde{x}^T Q \tilde{x} + \tilde{c}^T \tilde{x} \quad \text{where} \quad \tilde{c} \equiv (2Ql + c).
 \end{aligned}$$

So this substitution has no effect on the quadratic term of the objective function but does modify the linear term. In fact, a linear term could be created as a result of this substitution.

Next, consider the linear constraints. The effect of this shift on the inequality and equality constraints are similar. For the inequalities:

$$\begin{aligned} A_{ieq}x \leq b_{ieq} &\longrightarrow A_{ieq}(\tilde{x} + l) \leq b_{ieq} \\ &\longrightarrow A_{ieq}\tilde{x} \leq b_{ieq} - A_{ieq}l \\ &\longrightarrow A_{ieq}\tilde{x} \leq \widetilde{b_{ieq}} \quad \text{where} \quad \widetilde{b_{ieq}} \equiv b_{ieq} - A_{ieq}l. \end{aligned}$$

Similar work shows that $A_{eq}x = b_{eq} \longrightarrow A_{eq}\tilde{x} = \widetilde{b_{eq}}$ where $\widetilde{b_{eq}} \equiv b_{eq} - A_{eq}l$.

The effect of the shift is to zero out the lower bound of every component of \tilde{x} . This result should not be surprising since it was the original purpose behind this substitution. Analytically,

$$\begin{aligned} l \leq x \leq u &\longrightarrow l \leq (\tilde{x} + l) \leq u \\ &\longrightarrow l - l \leq \tilde{x} \leq u - l \\ &\longrightarrow 0 \leq \tilde{x} \leq \tilde{u} \quad \text{where} \quad \tilde{u} \equiv u - l. \end{aligned}$$

Finally, the effect of this shift on the orthogonality constraints must be examined. Assuming that $(i, j) \in \Phi$, then $x_i x_j = 0 \longrightarrow (\tilde{x}_i + l_i)(\tilde{x}_j + l_j) = 0$. Expanding this expression yields:

$$\tilde{x}_i \tilde{x}_j + \tilde{x}_i l_j + l_i \tilde{x}_j + l_i l_j = 0.$$

Notice that there are both quadratic and linear terms in this expression. For reasons discussed in Section 3.3, we require that $\tilde{x}_i \tilde{x}_j = 0$. Therefore the vector of lower bounds, l , is acceptable if and only if $l_i = l_j = 0 \forall (i, j) \in \Phi$. Checking that this additional requirement on l is satisfied can be implemented as a preprocessing step. In the context of Figure 2.1, every component of x must have a zero lower bound with the exception of x_2 .

Therefore after this first step (QPO) has the form:

$$\begin{aligned}
& \min_{\tilde{x}} && \tilde{x}^T Q \tilde{x} + \tilde{c}^T \tilde{x} \\
& \text{s.t.} && A_{ieq} \tilde{x} \leq \widetilde{b_{ieq}} \\
& && A_{eq} \tilde{x} = \widetilde{b_{eq}} \\
& && 0 \leq \tilde{x} \leq \tilde{u} \\
& && \tilde{x}_i \tilde{x}_j = 0 \quad \forall (i, j) \in \Phi, \text{ given acceptable } l.
\end{aligned}$$

2.2 Step 2: Introduce Slacks into Inequalities

Let s denote a vector of nonnegative slack variables, then

$$A_{ieq} \tilde{x} \leq \widetilde{b_{ieq}} \longrightarrow [A_{ieq} \quad \mathbf{I}] \begin{bmatrix} \tilde{x} \\ s \end{bmatrix} = \widetilde{b_{ieq}}, \quad s \geq 0.$$

Putting all the equalities together yields the following linear system $A\hat{x} = b$:

$$\begin{bmatrix} A_{ieq} & \mathbf{I} \\ A_{eq} & 0 \end{bmatrix} \begin{bmatrix} \tilde{x} \\ s \end{bmatrix} = \begin{bmatrix} \widetilde{b_{ieq}} \\ \widetilde{b_{eq}} \end{bmatrix}$$

Regarding the objective function, notice that Q and c can simply be padded with zeros to make the dimensionalities compatible. That is:

$$\begin{aligned}
\tilde{x}^T Q \tilde{x} + \tilde{c}^T \tilde{x} &= \begin{bmatrix} \tilde{x}^T & s \end{bmatrix} \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{x} \\ s \end{bmatrix} + \begin{bmatrix} \tilde{c}^T & 0 \end{bmatrix} \begin{bmatrix} \tilde{x} \\ s \end{bmatrix} \\
&\longrightarrow \hat{x}^T \hat{Q} \hat{x} + \hat{c}^T \hat{x}
\end{aligned}$$

One final point to notice is that every component of s is an orthogonally unconstrained variable. Thinking in terms of the graph Φ , adding these new slack variables corresponds to adding new nodes; however, no new edges are added. So the problem can now be re-expressed in terms of the augmented variable \hat{x} .

After Step 2, (QPO) has the following form:

$$\begin{aligned}
& \min_{\hat{x}} \quad \hat{x}^T \hat{Q} \hat{x} + \hat{c}^T \hat{x} \\
& \text{s.t.} \quad A \hat{x} = b \\
& \quad \quad 0 \leq \hat{x} \leq \tilde{u} \\
& \quad \quad \hat{x}_i \hat{x}_j = 0 \quad \forall (i, j) \in \Phi, \text{ given acceptable } l.
\end{aligned}$$

2.3 Step 3: Incorporate Linear Term in Quadratic Objective

This step is accomplished through the introduction of a new scalar variable, y , and an additional constraint, $y = 1$.

For the objective function, consider the following transformation:

$$\begin{aligned}
\hat{x}^T \hat{Q} \hat{x} + \hat{c}^T \hat{x} & \longrightarrow \begin{bmatrix} \hat{x}^T & y \end{bmatrix} \begin{bmatrix} \hat{Q} & \frac{\hat{c}}{2} \\ \frac{\hat{c}^T}{2} & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ y \end{bmatrix} \\
& = \begin{bmatrix} \hat{x}^T \hat{Q} + \frac{1}{2} y \hat{c}^T & \frac{1}{2} \hat{x}^T \hat{c} \end{bmatrix} \begin{bmatrix} \hat{x} \\ y \end{bmatrix} \\
& = \hat{x}^T \hat{Q} \hat{x} + \frac{1}{2} y \hat{c}^T \hat{x} + \frac{1}{2} \hat{x}^T \hat{c} y \\
& = \hat{x}^T \hat{Q} \hat{x} + \frac{1}{2} y \{ \hat{c}^T \hat{x} + \hat{x}^T \hat{c} \} \\
& = \hat{x}^T \hat{Q} \hat{x} + \frac{1}{2} y \{ 2 \hat{c}^T \hat{x} \} \\
& = \hat{x}^T \hat{Q} \hat{x} + y \hat{c}^T \hat{x}
\end{aligned}$$

However with the added constraint that $y = 1$, the objective function shown above has not been changed. Further, notice that in this construction the quadratic term remains symmetric. As an aside, we mention that introducing a nonzero “ y^2 term” simply places a constant into the objective function. Such a constant has no effect since we are minimizing this objective but can be useful for other reasons.

At this stage, the introduction of y would seem to require that the equality constraints be enlarged through the introduction of an appropriately-sized zero column vector.

Analytically:

$$A\hat{x} = b \longrightarrow [A \quad 0] \begin{bmatrix} \hat{x} \\ y \end{bmatrix} = b$$

However, it turns out that we can make double use of the variable y in a way that obviates the need for such padding, see Section 2.4. Finally just as with the slack variables, notice that introducing y does not impact the set of orthogonality constraints.

So dropping all accents at this point, (QPO) has been transformed into the following equivalent form:

$$\begin{aligned} \min_{x,y} \quad & \begin{bmatrix} x^T & y \end{bmatrix} \begin{bmatrix} Q & \frac{c}{2} \\ \frac{c^T}{2} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ \text{s.t.} \quad & Ax = b \\ & y = 1 \\ & 0 \leq x \leq u \\ & x_i x_j = 0 \quad \forall (i, j) \in \Phi, \text{ given acceptable } l. \end{aligned}$$

In a paper dealing with quadratic programming problems subject to bound and quadratic constraints, Ye first uses a scalar variable to obtain a strictly quadratic objective [48]. Here we go one step further by consolidating the incorporation of a linear term with the homogenization of any linear equalities through the use of a single variable. That is the topic of the next section.

2.4 Step 4: Homogenize Equalities

Reconsider $Ax = b$. Notice that this can be rewritten in the following form:

$$Ax = b \longrightarrow \begin{cases} Ax - by = 0 \\ y = 1 \end{cases}$$

where y is the same constrained scalar variable. Also, the sole inhomogeneous constraint involves only the scalar variable y .

Making this change, the final equivalent formulation of (QPO) is:

$$\begin{aligned}
 \min_{x,y} \quad & \begin{bmatrix} x^T & y \end{bmatrix} \begin{bmatrix} Q & \frac{c}{2} \\ \frac{c^T}{2} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
 \text{s.t.} \quad & Ax - by = 0 \\
 & y = 1 \\
 & 0 \leq x \leq u \\
 & x_i x_j = 0 \quad \forall (i, j) \in \Phi, \text{ given acceptable } l.
 \end{aligned}$$

This formulation can be relaxed as a semidefinite programming problem. It is the work of Chapter 3 to show exactly how to proceed. To set up that discussion, the following observations are made:

- Notice that there is a symmetric, strictly quadratic objective function.
- The constraints have been restated and now consist of a homogeneous set of linear equalities and only a single inhomogeneous constraint.
- The bound constraints are currently expressed in terms of each component of x . However, these individual bounds constitute implied constraints on pairs of components.
- Finally, the orthogonality constraints have maintained their required form.

CHAPTER 3

The Transition from (QPO) to (rSDP)

In the previous chapter, our standard form (QPO) was taken through a series of steps culminating in an equivalent formulation ready to be translated into the language and notation of a semidefinite programming problem. Before delving into specifics, a few general properties are presented.

3.1 Further Remarks on Semidefinite Programming

Consider the transformation $X = xx^T$. More explicitly,

$$X = xx^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1x_2 & \dots & x_1x_n \\ x_2x_1 & x_2^2 & \dots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & \dots & x_n^2 \end{bmatrix}$$

Thinking of $X_{ij} = x_i x_j$ establishes a correspondence between the entries in X and pairwise products of the components of x . This fundamental relationship between the vector variable x and the matrix variable X is exactly the type of correspondence that we need in order to proceed.

Notice that $X = xx^T$ implies that X is positive semidefinite. Consider the definition of a positive semidefinite matrix: $X \succeq 0 \Leftrightarrow v^T X v \geq 0 \forall v$. Here,

$$v^T X v = v^T (xx^T) v = (v^T x) (x^T v) = (x^T v)^T (x^T v) = |x^T v|^2 \geq 0 \quad \forall v.$$

So in fact $X = xx^T$ is a stronger condition than $X \succeq 0$. Described in words, $X = xx^T$ requires the solution be a rank-one matrix while $X \succeq 0$ only requires that the solution be positive semidefinite. As will be seen in Section 3.7, imposing the weaker of these two conditions is the only relaxation we will perform. This relaxation is fundamental to the polynomial complexity of semidefinite programming.

3.2 Implications of $x, y \longrightarrow Z$

Turning now to specifics, the general correspondence of the previous page can be written in terms of the variables of (QPO) to define the matrix variable, Z :

$$Z \equiv zz^T \equiv \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} x^T & y \end{bmatrix} = \begin{bmatrix} xx^T & xy \\ yx^T & y^2 \end{bmatrix} = \begin{bmatrix} X & xy \\ yx^T & y^2 \end{bmatrix}$$

The goal now is to translate the previous chapter's statements involving x and y into statements involving the matrix variable Z . The final step will be to weaken the rank-one restriction $Z = zz^T$ to simply $Z \succeq 0$. Before moving away from the block structure of Z , we point out that the definition of X on the previous page remains a useful shorthand for the first diagonal block of Z .

The easiest piece to translate is the objective. Since care was taken to construct a strictly quadratic objective, it is immediately expressible as a Frobenius product. Again, the Frobenius product is defined: $C \bullet Z = \text{trace}(CZ) = \sum_{i,j} C_{ij}Z_{ij}$. As an aside, this operation is the straightforward extension of the inner product of two vectors: $c \cdot z = \sum_i c_i z_i$.

Applying the definition of the Frobenius product means the following correspondence holds true:

$$\begin{bmatrix} x^T & y \end{bmatrix} \begin{bmatrix} Q & \frac{c}{2} \\ \frac{c^T}{2} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} Q & \frac{c}{2} \\ \frac{c^T}{2} & 0 \end{bmatrix} \bullet \begin{bmatrix} X & xy \\ yx^T & y^2 \end{bmatrix} \\ \longrightarrow C \bullet Z$$

where C is introduced for the objective matrix.

Next, all our constraints must be lifted and imposed on the matrix variable Z in this higher dimensional space. Subsequent sections progress in order of increasing complexity. All the constraints must have the form $A_k \bullet Z = b_k$ where A_k are matrices, b_k are scalars and the subscript k is simply used to index the constraints. As an aside, essentially all of the constraint matrices A_k are sparse. This is a fact that was fully exploited in the computational implementation.

3.3 Imposing Orthogonality Constraints

First notice that the orthogonality constraints can be directly imposed on the matrix variable. This is accomplished simply by constraining specific elements of X to be zero. Analytically, all that is necessary is to impose $X_{ij} = 0, \forall (i, j) \in \Phi$. With a simple manipulation, symmetry can be maintained in the constraint matrices as well.

Let $(i, j) \in \Phi$, then

$$\begin{aligned} x_i x_j = 0 &\implies 2x_i x_j = 0 \\ &\implies x_i x_j + x_j x_i = 0 \\ &\implies X_{ij} + X_{ji} = 0. \end{aligned}$$

Here the commutativity of scalar multiplication is at work. It is at this point that our reasons for requiring that $l_i = l_j = 0 \forall (i, j) \in \Phi$ become clear. If this were not true then we would not be able to impose orthogonality in this simple way.

As an aside, notice that this type of constraint matrix can be written as follows:

$$A_k = e_i e_j^T + e_j e_i^T$$

where e_i, e_j denote the i^{th} and j^{th} column unit vectors respectively. Such a matrix is super-sparse with ones located only in the (i, j) and (j, i) positions. This construction is mentioned here since it will reappear in a generalized form for the lifting procedure of Section 3.6.

3.4 Imposing Bound Constraints

Expanding on a remark from the end of Chapter 2, bounds on components of vector x can be easily translated into bounds on elements of the matrix X . The types of nontrivial constraints are: (1) upper bounds and (2) off-diagonal nonnegativity. Since the algebra is similar between the orthogonality and nonnegativity constraints, we will concentrate mostly on the issue of imposing upper bounds.

Assume that $(i, j) \notin \Phi$, $0 \leq x_i \leq u_i$ and $0 \leq x_j \leq u_j$. One immediate consequence is $x_i^2 \leq u_i^2$ and $x_j^2 \leq u_j^2$. These constraints can be restated in terms of the matrix variable: $X_{ii} \leq u_i^2$ and $X_{jj} \leq u_j^2$. So imposing appropriate bounds on diagonal elements of the matrix variable is straightforward.

Next, focus on the off-diagonal elements X_{ij} . We will show that imposing diagonal bounds, together with the requirement that X is symmetric positive semidefinite, suffices. In general for a matrix to be positive semidefinite, each principal submatrix must also be positive semidefinite. Consider the following 2×2 symmetric principal submatrix of X :

$$\begin{bmatrix} X_{ii} & X_{ij} \\ X_{ij} & X_{jj} \end{bmatrix} \succeq 0$$

Here, the requirement of positive semidefiniteness implies:

$$X_{ij}^2 \leq X_{ii}X_{jj}$$

Since $X_{ii}X_{jj} \leq u_i^2u_j^2$ is a consequence of intersecting the diagonal bounds then $X_{ij}^2 \leq u_i^2u_j^2$ or equivalently $X_{ij} \leq u_iu_j$ is actually being imposed off the diagonal. This is exactly the off-diagonal upper bound constraint that we require.

In terms of implementation, the inequalities presented above must be tightened to equality constraints in order to use SDPT3. A nonnegative slack variable can be introduced in a minimally expensive way to do this. This was possible with earlier versions of SDPT using a diagonal semidefinite block [44]. The most recent version of SDPT3 allows these slack variables to be placed within the linear block of an SQLP formulation [45]. In either case, the form of the resulting A_k matrix is again super-sparse. It is worth mentioning that other solution methods can directly handle inequality constraints, for example Helmberg's Spectral Bundle Method [23].

Before leaving this section, the important role of the nonnegativity constraints should be emphasized. Forcing the off-diagonal elements of Z to be nonnegative is a nontrivial type of constraint. It is not a consequence of $Z \succeq 0$.

3.5 Imposing $y = 1$

Looking at the structure of Z , one constraint that we could impose is $y^2 = 1$. This is simply expressible as the following matrix constraint:

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \bullet \begin{bmatrix} X & xy \\ yx^T & y^2 \end{bmatrix} = 1$$

Since $\{y^2 = 1\} \Rightarrow y = \pm 1$, it is not obvious that this constraint should be acceptable. We wanted to impose $y = +1$ and have been forced to allow the additional possibility that $y = -1$. However, we will illustrate that this allowance does not change the nature of the problem.

Consider the case $y = -1$. In particular, determine the impact on the homogenized constraints of (QPO):

$$\begin{aligned} Ax - by = 0 &\longrightarrow Ax - b(-1) = 0 \\ &\longrightarrow Ax + b = 0 \\ &\longrightarrow Ax = -b \\ &\longrightarrow A(-x) = b \end{aligned}$$

So it would seem that in this case we're finding $-x$, the antiparallel vector of what we would like. However, because of the form of the objective that we are minimizing this turns out not to be consequential. Consider the objective:

$$\begin{aligned} x^T Qx + yc^T x &\longrightarrow (-x)^T Q(-x) + yc^T(-x) \\ &\longrightarrow (-x)^T Q(-x) + (-1)c^T(-x) \quad \text{since } y = -1 \\ &\longrightarrow x^T Qx + c^T x \end{aligned}$$

The objective is unaffected. Therefore we are able to translate the single inhomogeneous constraint of (QPO) into a suitable matrix constraint on Z .

3.6 Symmetric Lifting of Homogenized Linear Equalities

The last and most complicated issue we must address is to how to impose the homogeneous linear equality constraints. The following sections detail the symmetric lifting procedure we have developed. Lifting is a general term for the type of procedure we present. Lifting (specifically for linear inequalities involving binary variables) is discussed by Goemans and Rendl [20].

Of interest is how a homogeneous linear constraint of the form $Az = 0$ can be enforced in a symmetric manner on the matrix variable Z . Consider $A \in \mathbb{R}^{m \times n}$. Analytically the lifting procedure we use can be expressed as follows:

$$\begin{aligned} Az &= 0 \\ Az(2z^T) &= 0(2z^T) \\ 2Azz^T &= [0 \quad 0 \quad \dots \quad 0] \\ (2Az)[z_1 \quad z_2 \quad \dots \quad z_n] &= [0 \quad 0 \quad \dots \quad 0] \end{aligned}$$

Notice that so long as $z \neq 0$ then any solution to $Azz^T = 0$ also satisfies $Az = 0$. With this requirement, lifting the original linear system $Az = 0$ does not weaken it. The preceding matrix equality corresponds to the following homogeneous system:

$$\left\{ \begin{array}{l} (2Az) z_1 = 0 \\ (2Az) z_2 = 0 \\ \vdots \\ (2Az) z_n = 0 \end{array} \right\}$$

Letting $a_l = l^{th}$ row of A then for $j = 1, \dots, n$ and $l = 1, \dots, m$ we have:

$$\left(2 \sum_{i=1}^n a_{li} z_i \right) z_j = 0$$

$$\sum_{i=1}^n a_{li} \{2z_i z_j\} = 0$$

By the commutativity of scalar multiplication, this becomes:

$$\sum_{i=1}^n a_{li} \{z_i z_j + z_j z_i\} = 0$$

Or in terms of our matrix variable,

$$\sum_{i=1}^n a_{li} \{Z_{ij} + Z_{ji}\} = 0$$

Incorporating orthogonality, the entire system can be represented as:

$$\sum_{\substack{i=1 \\ (i,j) \notin \Phi}}^n a_{li} \{Z_{ij} + Z_{ji}\} = 0 \quad \text{for } j = 1, \dots, n \text{ and } l = 1, \dots, m.$$

Notice that since $A \in \mathbb{R}^{m \times n}$, there are mn matrix constraints generated by this procedure. However, the structure of the corresponding A_k constraint matrices may not be obvious in this construction. So for illustrative purposes, a small example is presented next. This is followed in Section 3.6.2 by an alternative analytic construction which more clearly shows the structure of the lifted constraint matrices.

3.6.1 Lifting Example

Consider the simple example of $A \in \mathbb{R}^{1 \times 3}$. More explicitly, $A \equiv [a_{11} \ a_{12} \ a_{13}]$. Also, for now assume $\Phi = \emptyset$. Then $Az = 0$ is simply $a_{11}z_1 + a_{12}z_2 + a_{13}z_3 = 0$. Rewriting in vector form:

$$[a_{11} \ a_{12} \ a_{13}] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = 0$$

Now right-multiply by $2z^T$ and expand the outer product:

$$2[a_{11} \ a_{12} \ a_{13}] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} [z_1 \ z_2 \ z_3] = 2 \cdot 0 \cdot [z_1 \ z_2 \ z_3]$$

\Updownarrow

$$2[a_{11} \ a_{12} \ a_{13}] \begin{bmatrix} z_1^2 & z_1 z_2 & z_1 z_3 \\ z_2 z_1 & z_2^2 & z_2 z_3 \\ z_3 z_1 & z_3 z_2 & z_3^2 \end{bmatrix} = [0 \ 0 \ 0]$$

That matrix equality corresponds component-by-component to the following homogeneous system:

$$\begin{cases} 2a_{11}z_1^2 + a_{12}(2z_2z_1) + a_{13}(2z_3z_1) = 0 \\ a_{11}(2z_1z_2) + 2a_{12}z_2^2 + a_{13}(2z_3z_2) = 0 \\ a_{11}(2z_1z_3) + a_{12}(2z_2z_3) + 2a_{13}z_3^2 = 0 \end{cases}$$

\Updownarrow

$$\begin{cases} 2a_{11}z_1^2 + a_{12}(z_2z_1 + z_1z_2) + a_{13}(z_3z_1 + z_1z_3) = 0 \\ a_{11}(z_1z_2 + z_2z_1) + 2a_{12}z_2^2 + a_{13}(z_3z_2 + z_2z_3) = 0 \\ a_{11}(z_1z_3 + z_3z_1) + a_{12}(z_2z_3 + z_3z_2) + 2a_{13}z_3^2 = 0 \end{cases}$$

Next, expand all terms and number the equalities using the index j .

$$\begin{cases} 2a_{11}z_1^2 + a_{12}z_2z_1 + a_{12}z_1z_2 + a_{13}z_3z_1 + a_{13}z_1z_3 = 0 & (j=1) \\ a_{11}z_1z_2 + a_{11}z_2z_1 + 2a_{12}z_2^2 + a_{13}z_3z_2 + a_{13}z_2z_3 = 0 & (j=2) \\ a_{11}z_1z_3 + a_{11}z_3z_1 + a_{12}z_2z_3 + a_{12}z_3z_2 + 2a_{13}z_3^2 = 0 & (j=3) \end{cases}$$

Switching to the Frobenius product, notice that for $j=1$:

$$\begin{bmatrix} 2a_{11} & a_{12} & a_{13} \\ a_{12} & 0 & 0 \\ a_{13} & 0 & 0 \end{bmatrix} \bullet \begin{bmatrix} z_1^2 & z_1z_2 & z_1z_3 \\ z_2z_1 & z_2^2 & z_2z_3 \\ z_3z_1 & z_3z_2 & z_3^2 \end{bmatrix} = 0$$

Next for $j=2$:

$$\begin{bmatrix} 0 & a_{11} & 0 \\ a_{11} & 2a_{12} & a_{13} \\ 0 & a_{13} & 0 \end{bmatrix} \bullet \begin{bmatrix} z_1^2 & z_1z_2 & z_1z_3 \\ z_2z_1 & z_2^2 & z_2z_3 \\ z_3z_1 & z_3z_2 & z_3^2 \end{bmatrix} = 0$$

Finally for $j=3$:

$$\begin{bmatrix} 0 & 0 & a_{11} \\ 0 & 0 & a_{12} \\ a_{11} & a_{12} & 2a_{13} \end{bmatrix} \bullet \begin{bmatrix} z_1^2 & z_1z_2 & z_1z_3 \\ z_2z_1 & z_2^2 & z_2z_3 \\ z_3z_1 & z_3z_2 & z_3^2 \end{bmatrix} = 0$$

These constraints are clearly of the form $A_k \bullet Z = b_k$. In fact, $b_k = 0$ for all lifted constraints. It is also much easier to appreciate the special sparsity pattern of these lifted A_k constraint matrices. That sparsity pattern motivates the more compact construction presented in Section 3.6.2.

As a final point, suppose $z_1 z_3 = 0$ was required. That is to say $(1, 3) \in \Phi$. That orthogonality could be imposed by:

1. zeroing out the $(1, 3)$ and $(3, 1)$ positions in every A_k matrix and
2. constructing the appropriate matrix constraint to force $z_1 z_3 = 0$.

3.6.2 Alternate Analytic Construction for $Az = 0$

The characterization of the lifting procedure given in Section 3.6 shows a logical progression from start to finish. However, it is not compact notationally. Inspired by the structure of the lifted matrices, the lifting procedure can be more simply expressed as a sum of two rank-one operations.

Let $a_l = l^{th}$ row of A written as a row vector and let $e_j = j^{th}$ unit vector written as a column vector. It will be shown that $\{a_l^T e_j^T + e_j a_l\} \bullet Z = 0$ is our constraint. Starting with the definition of the Frobenius product, consider the ansatz:

$$\begin{aligned}
 \{a_l^T e_j^T + e_j a_l\} \bullet Z &= \text{trace} \left(\{a_l^T e_j^T + e_j a_l\} Z \right) \\
 &= \text{trace} \left(\{a_l^T e_j^T + e_j a_l\} z z^T \right) \\
 &= z^T \{a_l^T e_j^T + e_j a_l\} z \\
 &= z^T \{a_l^T (e_j^T z) + e_j a_l z\} \\
 &= z^T \{a_l^T z_j + e_j a_l z\} \\
 &= z^T a_l^T z_j + (z^T e_j) a_l z \\
 &= z^T a_l^T z_j + z_j a_l z \\
 &= (a_l z)^T z_j + z_j a_l z \\
 &= 2z_j(a_l z) = 0 \text{ which is exactly our constraint.}
 \end{aligned}$$

So a compact notation to represent all the necessary lifted A_k constraints is:

$$A_k = a_l^T e_j^T + e_j a_l \text{ for } j = 1, \dots, n \text{ and } l = 1, \dots, m.$$

Notice how this construction matches that found in Section 3.3 where $e_i^T \longrightarrow a_l$.

3.7 Relaxing (QPO) into (rSDP)

The preceeding work has shown how to restate (QPO) in the following form:

$$\begin{aligned} \min_Z \quad & C \bullet Z \\ \text{s.t.} \quad & A_k \bullet Z = b_k \quad \text{for } k = 1, \dots, K \\ & Z = zz^T \end{aligned}$$

We now relax the rank-one constraint on Z and obtain (rSDP):

$$\begin{aligned} \text{(rSDP)} \quad & \min_Z \quad C \bullet Z \\ \text{s.t.} \quad & A_k \bullet Z = b_k \quad \text{for } k = 1, \dots, K \\ & Z \succeq 0 \end{aligned}$$

Our solution strategy for (QPO) is to solve this semidefinite relaxation (rSDP) as the first step in an enumerative algorithm. The complete explanation of this strategy can be found in Chapter 4.

3.8 Relating the Sizes of (QPO) and (rSDP)

Having illustrated the theoretical correspondence between (QPO) and (rSDP), one practical question that should be answered surrounds the relationship between the sizes of (QPO) and (rSDP). As you would expect, there is a well defined relationship between the dimensionalities of each instance of (QPO) and the dimensionalities of its corresponding (rSDP). Table 3.1 on the following page presents such a calculation for an instance of (QPO) relevant to the financial application of Chapter 5. The result of this analysis is that for $x \in \mathbb{R}^n$, $K \sim O(n^2)$ constraints must be imposed on (rSDP). Using SDPT, this growth in the number of constraints places a practical limit on tractable portfolio size of around 30 securities.

Object	Description	Dimensionality
(QPO)		
x	vector variable	$n \times 1$
Q	quadratic objective	$n \times n$
c	linear objective	$n \times 1$
A_{ieq}	linear inequality coefficients	$m_{ieq} \times n$
b_{ieq}	linear inequality rhs.	$m_{ieq} \times 1$
A_{eq}	linear equality coefficients	$m_{eq} \times n$
b_{eq}	linear equality rhs.	$m_{eq} \times 1$
l, u	lower and upper bounds	$n \times 1$
n_{op}	number of orthogonal pairs	$n/2$
Total Number of Variables		$(n + m_{ieq} + 1) \equiv \tilde{n}$
Total Number of Equality Constraints		$(m_{ieq} + m_{eq} + 1) \equiv \tilde{m}$
(rSDP)		
Z	matrix variable	$\tilde{n} \times \tilde{n}$
C	objective matrix	$\tilde{n} \times \tilde{n}$
A_k	constraint matrices (size)	$\tilde{n} \times \tilde{n}$
A_k	constraint matrices (source)	
	$y^2 = 1$	1
	orthogonality	n_{op}
	diagonal upper bounds	$n + 1$
	off-diagonal nonnegativity	$O(\tilde{n}^2/2)$
	lifting $Az = 0$	$\tilde{m}\tilde{n}$
Total Number of SDP constraints		$O(\tilde{n}^2 + \tilde{m}\tilde{n})$
Note: This is the length of b_k , the SDP constraint rhs. vector.		

Table 3.1: For a (QPO) representing the portfolio rebalancing problem involving n_s securities, then $n = 2n_s$ and $n_{op} = n_s$. In this application, there is a single inequality and one or two equality constraints.

CHAPTER 4

(DOP) Solution Strategy and Test Problems

At this stage, we have explained our construction of a semidefinite relaxation for (QPO). The focus of this chapter now turns to our solution strategy; specifically, how that relaxation is used to solve certain instances of (QPO). Before continuing, though many of the arguments generalize for other classes of orthogonality conditions, here we will only be considering the (DOP) case.

4.1 Special Structure of (DOP) Orthogonality

Refreshing the reader, (DOP) is the case where orthogonality relationships exist only within disjoint pairs of variables. Here when $x_i x_j = 0$ then, for the definitive cases, either $x_i > 0$ or $x_j > 0$ not both. This must be true without having to consider anything else in the problem. In other words, for the definitive cases exactly one member of an orthogonal pair of variables must be nonzero and the other member must be zero.

As will be seen in Chapter 5, this is exactly the type of combinatorial structure that is present in the portfolio rebalancing problem. So unlike a general set of orthogonality relationships, the combinatorial aspect of (DOP) is simple to express and takes precedence.

4.2 Remarks on Terminology

Since we will be using terminology that developed and adapted over the time-frame of this research, it seems appropriate to take a moment and define that terminology for a wider audience in hopes of heading off confusion.

The first piece of terminology to be explained is that of “activity.” Our heuristic decision-making procedure makes determinations regarding activity. By this we mean that a determination is made regarding which member of an orthogonal pair of variables should be allowed to be nonzero. Stated in a complementary fashion, a determination is being made that one member of an orthogonal pair should be set

to zero. These determinations are made only if our estimate for the optimal solution satisfies a specific criteria. The estimate we use is discussed in Section 4.6.1; the criteria is discussed in Section 4.6.2. Failing to satisfy that criteria means no determination is made. Without a determination of activity, then both nonzero possibilities for that orthogonal pair are considered.

We will refer to the nonzero member of an orthogonal pair as being the “active” member of the pair. Collectively, the term “active set” refers to the set of variables that are allowed to be nonzero. This usage is not to be confused with the term active set of constraints that refers to the set of inequality constraints satisfied with equality. Of course given the zero lower bound assumed for all members of orthogonal pairs, there is a clear connection between these two ideas.

The second piece of terminology that needs to be addressed is the issue of “degeneracy” and “degenerate decisions.” We will postpone a full discussion of this idea until Chapter 6 since it is more easily illustrated after the portfolio rebalancing problem has been introduced. However in terms of a working definition, a degenerate decision is one where the solution returned as optimal contains an active variable that is zero.

4.3 Simple Example: (QPO) with (DOP) Orthogonality

In Section 1.3.1, we first presented (DOP) orthogonality. Before describing our solution strategy, it is important that the exact problem we are considering be clearly illustrated. In our terminology, we are considering instances of (QPO) with (DOP) orthogonality. These are quadratic programming problems where in addition to any linear constraints and bounds, pairs of variables can be orthogonally constrained.

Concentrating on the original work of this thesis, consider a simple example where $x \in \mathbb{R}_+^n$ is subject to a set of linear equalities and among the set of n variables there are two disjoint orthogonal pairs.

$$\begin{aligned} \min_x \quad & x^T Q x + c^T x \\ & Ax = b \\ & x_i x_j = 0 \\ & x_k x_l = 0 \end{aligned}$$

This problem breaks into consideration of the following $2^2 = 4$ distinct subproblems, each corresponding to a possible active set.

Possibility (1): x_i and x_k are active

$$QPS_1 = \min_x \left\{ x^T Q x + c^T x \quad s.t. \quad Ax = b, \quad x_j = 0, \quad x_l = 0 \right\}$$

Possibility (2): x_i and x_l are active

$$QPS_2 = \min_x \left\{ x^T Q x + c^T x \quad s.t. \quad Ax = b, \quad x_j = 0, \quad x_k = 0 \right\}$$

Possibility (3): x_j and x_k are active

$$QPS_3 = \min_x \left\{ x^T Q x + c^T x \quad s.t. \quad Ax = b, \quad x_i = 0, \quad x_l = 0 \right\}$$

Possibility (4): x_j and x_l are active

$$QPS_4 = \min_x \left\{ x^T Q x + c^T x \quad s.t. \quad Ax = b, \quad x_i = 0, \quad x_k = 0 \right\}$$

The optimal objective value for the full problem is simply:

$$\min \{QPS_1, QPS_2, QPS_3, QPS_4\}.$$

This example is didactic but it clearly illustrates several important points. First, assuming Q is positive semidefinite then each possible active set corresponds to a convex QP-subproblem. Collectively, these QP-subproblems exhaust all possibilities. The optimal solution vector is simply identified as the solution vector for the optimal subproblem.

In terms of implementation, the dimensionality of the subproblems is reduced since inactive variables can be dropped. Also, though we do not currently take computational advantage, these subproblems can be solved in parallel.

4.3.1 The Naïve Approach for Solution

The previous example was small enough that we were able to find the optimal solution vector by simply exhausting all possibilities. This naïve approach solves the problem in a straightforward manner; there is no need for any outside knowledge of the optimal active set. Let us generalize this naïve approach.

Consider an instance of (QPO) with (DOP) orthogonality where there are n_{op} orthogonal pairs of variables in the problem. For this naïve approach, we must generate and solve all possible QP-subproblems, label these QPS_i . The optimal solution is identified with the QP-subproblem having the minimal objective value, that is $\min\{QPS_i\}$. The optimal solution vector is simply the corresponding solution vector $\vec{x}_{\hat{i}}$ where $\hat{i} = \operatorname{argmin}\{QPS_i\}$.

However, the number of possibilities grows exponentially with the number of orthogonal pairs involved in the problem. So for n_{op} orthogonal pairs then $2^{n_{op}}$ convex QP-subproblems would have to be considered in order to be guaranteed of finding the optimal solution. This approach may work perfectly well when $n_{op} = 2$ and there are only 4 QP-subproblems to consider. It will certainly not work much beyond that. (Aside: $2^{10} = 1024$ subproblems, $2^{20} \sim 1$ million subproblems, etc.)

So if this is not a practical procedure, why did we present it?

1. It clearly illustrates the curse of dimensionality for NP-hard problems.
2. If n_{op} is sufficiently small then this approach can provide a means to check the solution strategy we present in Section 4.4. One of the datasets used in Chapter 5 was selected for this reason.

4.3.2 Usefulness of Accurate Active Set Information

Since the naïve approach will not work, reconsider solving our problem. Suppose there were an oracle which provided complete information about the optimal active set. As the reader probably expects, such information would make solution trivial. Knowing the optimal active set would select from all possibilities only a single, ordinary, convex QP which would then have to be solved in order to identify the optimal solution. To ground this insight in our simple example, suppose the oracle said:

The optimal active set includes both x_j and x_k .

Then you would only bother solving the 3rd QP-subproblem and ignore the others.

Such an oracle does not exist. However, even partial information about the optimal active set would clearly be valuable. So the outstanding issues are: (1) How can that type of information be deduced? and (2) How accurate is it?

4.4 Our Solution Strategy: (rSDP) + QP-subproblems

This leads our discussion directly to the role played by (rSDP). While not an oracle, (rSDP) gives us a chance to deduce something about which variables are members of the optimal active set. We let the solution matrix of (rSDP) provide a rationale for reducing the size of the search space. In other words, we use (rSDP) to eliminate a large number of presumably sub-optimal QP-subproblems from consideration. Though the number of subproblems is still worst-case exponential, for each orthogonal pair where a determination of activity is made, the number of subproblems that must be considered is cut by $\frac{1}{2}$. If enough decisions are made, the number of subproblems that must be considered remains tractable. A more extensive analysis of the number of QP-subproblems considered by our strategy is presented in Chapter 6.

Stating our strategy more formally, we solve (rSDP) as the first stage of an enumerative algorithm. The solution matrix for (rSDP) provides an estimate for the optimal solution vector x . Based on a criteria, that estimate for x is used to deduce information about membership in the optimal active set. Such information allows

us to identify the most promising QP-subproblems which are then solved and the best observed solution is identified.

As a secondary and corrective step, the best identified solution vector is recursively examined for degeneracy. If degeneracy is present then based on the geometry of the feasible region, additional QP-subproblems corresponding to the neighboring faces are also solved. The final result of this process is returned as optimal. Figure 4.1 presents this strategy using a flowchart.

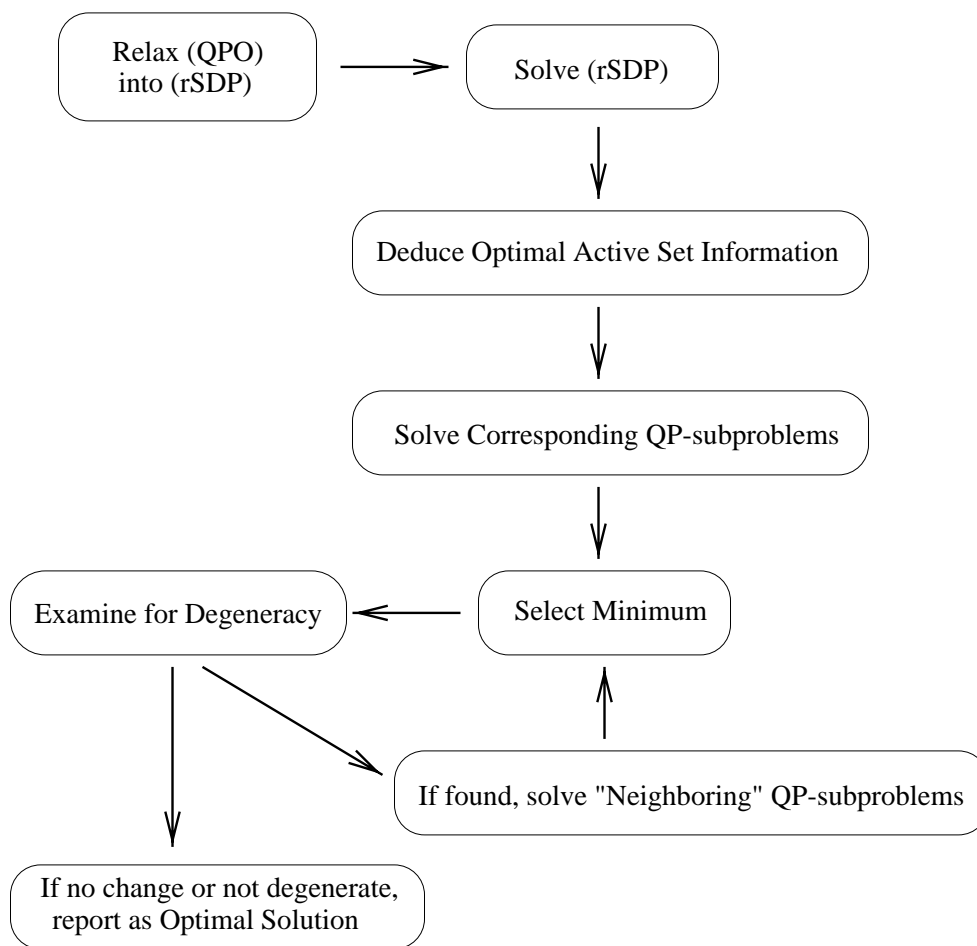


Figure 4.1: Flowchart illustrating the solution strategy used.

4.5 Description as an Enumerative Algorithm

As was mentioned previously, our approach can be classified as an enumerative algorithm. Here a loose analogy can be made to other enumerative algorithms, such as branch-and-bound or the simplex method in linear programming. For example, the simplex method provides a deterministic procedure for identifying which set of constraints are active at optimality. From some starting guess at the optimal set of active constraints, the simplex method moves to improved adjacent basic feasible solutions until the optimal set is found.

As is the case with our strategy, the worst-case performance of simplex remains exponential. However, despite the fact that simplex does not have guaranteed polynomial-time performance like interior-point or ellipsoidal schemes, the average performance of simplex is sufficiently good that it is used to solve a wide range of linear programming problems.

One metaphor for enumerative algorithms is that of a tree. Speaking in those terms, the idea behind our strategy is that solving (rSDP) gives us a way to immediately go deep into the tree. We hope to be able to correctly eliminate a large number of leaves from consideration. With the recursive degeneracy check, then additional leaves can be considered as justified. Figure 4.2 illustrates our strategy using this tree metaphor.

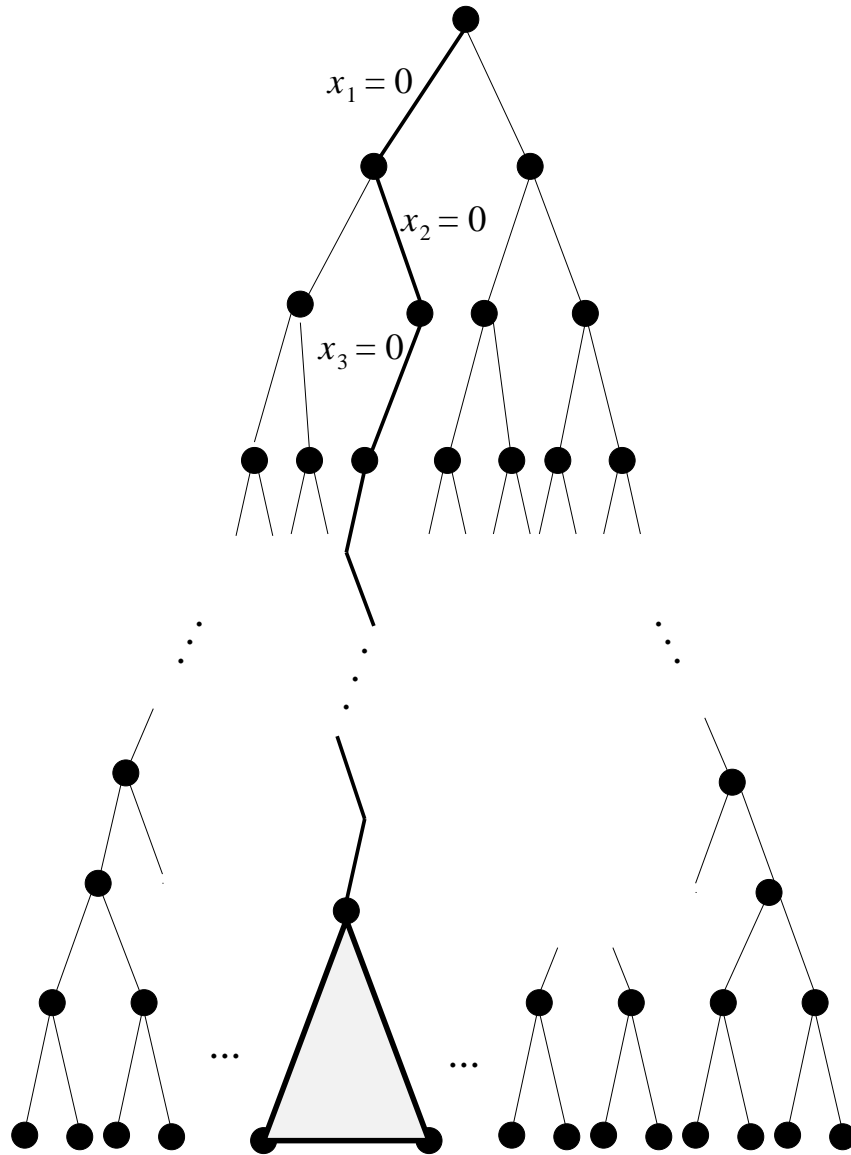


Figure 4.2: The figure uses the tree metaphor for our solution strategy. By solving (rSDP) and determining members of the optimal active set, only the small shaded set of QP-subproblems need to be initially considered. Here the bold lines illustrate where activity decisions have been made within the tree.

4.6 Deduction of Optimal Active Set Information

Having laid out the solution strategy, we now discuss the decision-making process that we use to determine active set information. We require an estimate of the optimal solution vector x , denoted x^* , as an input and then apply a criteria in order to determine which variables should be considered active and for which variables no determination of activity is possible. We present the computational procedure used to generate that estimate followed by a discussion of the decision criteria.

4.6.1 Estimating x^* from Z^*

As was discussed in Section 3.2, though our construction was based on identifying Z with zz^T , (rSDP) only requires the weaker condition that $Z \succeq 0$. In terms of implementation, the various SDP solvers are all iterative procedures and the optimal solution -or any intermediate solution- is only guaranteed to be positive semidefinite and not rank-one. That is in fact what we observe. Even solving (rSDP) to optimality, the solution matrix is not found to be strictly rank-one. As such, the optimal solution of (rSDP) is not immediately the optimal solution vector or even a feasible point for (QPO). In short, solving (rSDP) all the way to optimality does not eliminate the need to generate and solve some subset of QP-subproblems.

Since we are using (rSDP) as the first stage in an enumerative algorithm, it is not necessarily the most productive investment of computation to solve (rSDP) all the way to optimality. Stopping at some intermediate stage can still provide enough information to make valid activity decisions. The computational results presented in Section 5.7 exemplify the empirical accuracy of this remark.

Returning to our procedure for the estimate x^* , let Z^* denote a solution matrix returned for (rSDP). This solution matrix may be an intermediate iterate or could be the solution returned as optimal from the SDP solver. Listing its properties: Z^* is a symmetric, positive semidefinite matrix. As such, the spectral decomposition can be applied. Analytically:

$$Z^* = \sum_{i=1}^{\tilde{n}} \lambda_i \xi_i \xi_i^T$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_i \geq \dots \geq \lambda_{\tilde{n}}$ are the eigenvalues of Z^* listed in order of

decreasing magnitude. The eigenvector associated with each λ_i is denoted by ξ_i . Relevant dimensionalities are \tilde{n} for Z and n for x , refer to Table 3.1 if necessary.

Let $Z^1 = \lambda_1 \xi_1 \xi_1^T$. This is certainly a rank-one matrix; in fact it has the attractive property of being the “closest” rank-one matrix to Z^* w.r.t. $\|\cdot\|_2$. Given $Z_{ij} \geq 0 \forall i, j$ then $\xi_1 \geq 0$ follows from an analysis of the power method. Background on the power method, spectral decomposition or singular value decomposition more generally can be found in the following texts [15, 22]. As an aside, the second eigenvalue, λ_2 , measures how close the full matrix is to being rank-one, analytically: $\|Z^* - Z^1\|_2 = \lambda_2$. Computationally, λ_2 was found to be a solver-independent measure of progress towards a rank-one solution matrix.

Our estimate x^* comes directly from this rank-one matrix Z^1 . The following analytic construction identifies the relationship that exists between x^* and Z^1 :

$$\begin{aligned} Z^1 &= \lambda_1 \xi_1 \xi_1^T \\ &= \begin{pmatrix} \sqrt{\lambda_1} \xi_1 \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} \xi_1^T \end{pmatrix} \\ &\equiv \begin{bmatrix} x^* \\ y^* \end{bmatrix} \begin{bmatrix} x^{*T} & y^* \end{bmatrix} \end{aligned}$$

We compute x^* after using the spectral decomposition to find the closest rank-one matrix, Z^1 , to a solution of (rSDP). For clarity, we summarize this procedure:

Step 1: Solve (rSDP) until a user-specified stopping criteria is satisfied. Several examples of valid stopping criteria are: a fixed number of iterations, a time limit, a threshold on the second eigenvalue λ_2 , or a relative precision optimality criteria.

Step 2: Take the solution returned from (rSDP) and apply the spectral decomposition to find the closest rank-one matrix, $Z^1 = \lambda_1 \xi_1 \xi_1^T$.

Step 3: Estimate $x^* = \sqrt{\lambda_1} \xi_1(1:n)$. Here the notation $\xi_1(1:n)$ refers to the first n elements of ξ_1 .

Step 4: Use x^* to make activity decisions, see Section 4.6.2 on the following page for details of our approach.

To conclude this section, estimating x^* via the spectral decomposition is numerically stable and makes fullest use of all information contained in the (rSDP) solution matrix. Other estimates are possible, for example $x^* = \sqrt{\text{diag}(X^*)}$ where X^* is the first block of Z^* . However, these are only useful if (rSDP) is solved to optimality. As mentioned previously, this is not necessarily the most productive investment of computation.

4.6.2 Decision Criteria for Active Set Information

For an orthogonal pair (x_i, x_j) , a determination of activity means deciding which member of the pair should be allowed to be nonzero. From the previous procedure we have estimates (x_i^*, x_j^*) for both members of each orthogonal pair of variables. In order to make a decision, we look to see which variable was pushed farther away from zero. Or stated in a more useful way: we look to see which variable was pushed closer to its upper bound in percentage terms. Figure 4.3 on the following page contains three panels that illustrate features of the decision-making criteria for a particular orthogonal pair (i, j) .

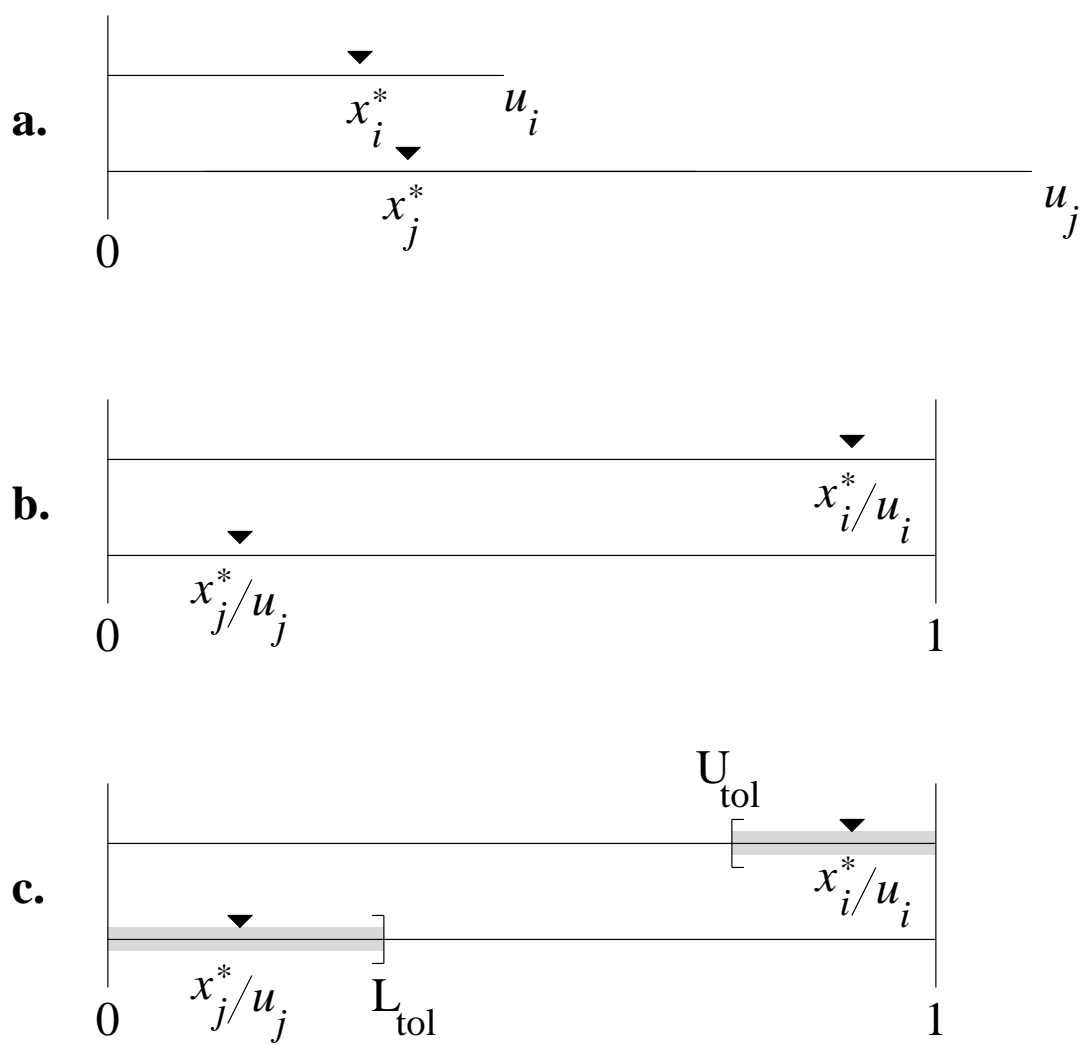


Figure 4.3: Illustration of Percent Upper Bound Criteria.

The differing lengths of the two horizontal lines in Figure 4.3.a. illustrate that while both lower bounds of an orthogonal pair must be zero, they can have differing upper bounds (u_i, u_j) . This will be exactly the case for the portfolio rebalancing application where the upper bounds on orthogonal buying and selling variables are different. To correct for that asymmetry, the percent of upper bound criteria was developed.

Figure 4.3.b. shows two horizontal lines that are now both of unit length. This highlights the fact that both estimates are being compared on the same basis; namely, as a percent of their upper bounds. Using this criteria, x_i^* now appears more significant. However, a decision is made only if allowed by two user-specified parameters.

The final stage is to compare the estimates (x_i^*, x_j^*) against each other and against two tolerances. These tolerances, L_{tol} and U_{tol} , are user-specified parameters that control the selectivity of the activity decision. Figure 4.3.c. provides that context. L_{tol} is a lower tolerance and U_{tol} is an upper tolerance. Only if one estimate for each member of an orthogonal pair lies in each shaded interval is a decision of activity made. Looking at this figure, then x_i is made active. Analytically, our decision-making criteria can be expressed as follows:

If $\left\{ \frac{x_i^*}{u_i} \geq U_{tol} \text{ and } \frac{x_j^*}{u_j} \leq L_{tol} \right\}$ then x_i should be considered active.

If $\left\{ \frac{x_i^*}{u_i} \leq L_{tol} \text{ and } \frac{x_j^*}{u_j} \geq U_{tol} \right\}$ then x_j should be considered active.

An obvious point, yet deserving mention, is that the criteria used most certainly affects the decision made. Applying a criteria such as $\max(x_i^*, x_j^*)$ as seen in Figure 4.3.a. would make the opposing decision that x_j^* should be considered active. So other criteria are certainly possible; one idea for a heuristic that is perhaps better suited for larger problems is described in Chapter 7.

Finally, we again emphasize that a determination of activity does not always occur for every orthogonal pair. In such an event, we simply must consider QP-subproblems corresponding to both possible active sets. So the number of activity decisions made is inversely related to the number of QP-subproblems that are considered.

4.7 Creation of Test Problems and Results

Before turning to a specific application in Chapter 5, we consider two types of test problems. These test problems were constructed to determine the effectiveness of our decision-making heuristic. Please note that for these test problems, a priori we know an analytic procedure to find the optimal solution. So in some sense, these test problems are easy. However when presented as instances of (QPO) and our solution strategy is applied, they are nontrivial to solve.

4.7.1 Linear Programming Problems

This is the first type of test problem we constructed. The KKT conditions associated with a linear program provide an example of (QPO) having our (DOP) orthogonality structure. There the orthogonal pairs are each primal variable and its corresponding dual slack. The orthogonality relationship also has a special name, complementary slackness. To construct instances of this test problem, consider the following procedure:

First construct u as a random vector of upper bounds. Next construct random vectors x, y, z subject to those bounds and a random matrix A . When constructing x and z , maintain complementary slackness. Let $b = Ax$ and $c = A^T y + z$. The KKT conditions with two-norm minimization provided as the objective give rise to the following (QPO):

$$\begin{aligned} \min_{x,y,z} & \begin{bmatrix} x^T & y^T & z^T \end{bmatrix} \mathbf{I} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ \text{s.t.} & \begin{bmatrix} A & 0 & 0 \\ 0 & A^T & \mathbf{I} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix} \\ & 0 \leq x, y, z \leq u \\ & \text{and } x_i z_i = 0 \quad \forall i. \end{aligned}$$

Given how the problem was created and the objective that was chosen, there is no question regarding the optimal solution. However, notice that the primal

feasibility and dual feasibility constraints are separable. In other words, there is no constraint linking both members x_i and z_i of any orthogonal pair.

Without such a linking constraint, we found empirically that although (rSDP) was successful in providing a tight lower bound on the optimal objective value, no decisions were being made for any orthogonal pair. Therefore, every possible subproblem would have to be considered in order to find the optimal solution vector. (DOP) orthogonality is present in this first type of test problem but there is no constraint where that orthogonality must act. The presence of a linking constraint became a requirement in order to test our heuristic in a meaningful way. The second type of test problem satisfies this requirement.

4.7.2 Random Symmetric Quadratic Test Problem

This test problem also involves a purely quadratic objective. In fact, the objective matrix is a rank-one creation. Since any orthogonally uninvolved variables do not affect the number of possible QP-subproblems, we consider only orthogonal pairs. Use the following procedure:

Specify some number of pairs of orthogonal variables, n_{op} . Generate a positive random vector, q , and let $Q = qq^T$. Generate a random number b_{ieq} . Consider the following instance of (QPO):

$$\begin{aligned} \min_x \quad & -x^T Q x \\ \text{s.t.} \quad & \sum_{i=1}^n x_i \leq b_{ieq} \\ & 0 \leq x \leq 1 \quad \forall i \\ & x_i x_{(i+n_{op})} = 0 \quad \text{for } i = 1, \dots, n_{op}. \end{aligned}$$

Notice that the objective function can be stated as: $-x^T qq^T x \rightarrow -(q^T x)^2$.

Without orthogonality, this is a continuous knapsack problem. The following greedy algorithm solves that problem. First, sort q to identify the largest entries. The optimal solution is to set $x_i = 1$ for the indices corresponding to the set of $\lfloor b_{ieq} \rfloor$ largest entries in q . For the $\lceil b_{ieq} \rceil^{\text{st}}$ largest entry of q , set $x_i = b_{ieq} - \lfloor b_{ieq} \rfloor$. So without orthogonality, the greedy algorithm is basically to select the largest entries q_i wherever they occur within the vector q .

To solve our test problem with orthogonality constraints, the same greedy approach works with only slight modification. Reconsider the problem with the addition of orthogonality. Now, the position of the largest entries of q becomes important. If entry q_i has been selected then $q_{(i \pm n_{op})}$ cannot also be selected. However, this is a simple operation to implement. So the optimal solution can again be found in greedy fashion, by sorting and selecting from among the largest entries of q .

Figure 4.4 on the following page contains results for a range of problem sizes of this second type of test problem. This plot demonstrates the variability in the performance of the heuristic for making active set decisions. While the worst case did occur in which no QP-subproblems could be eliminated from consideration, there were also a significant number of trials where a substantial fraction of subproblems were correctly eliminated from consideration.

In Figure 4.4, the SDPT solver was allowed to solve each instance of (rSDP) until optimality was reached as determined by its default relative precision criteria [44]. Throughout a conservative set of tolerances ($L_{tol} = 10\%$ and $U_{tol} = 90\%$) was used to make activity decisions during these random trials.

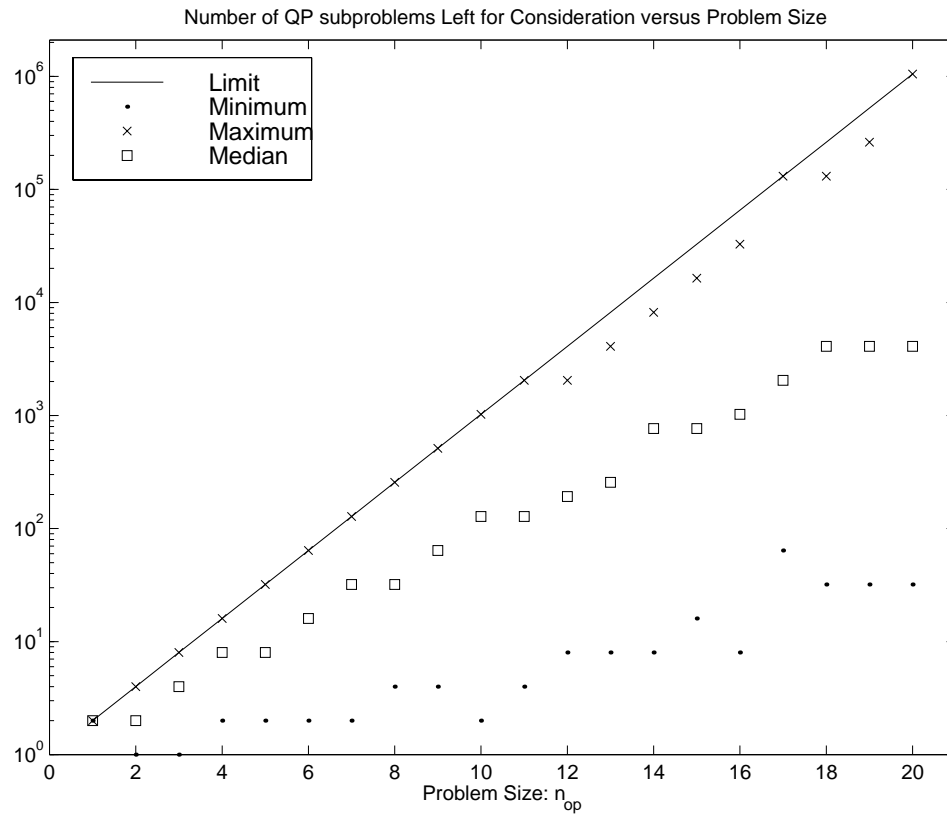


Figure 4.4: The line on this semilog plot shows that the number of potential QP-subproblems grows exponentially. One hundred trials were conducted for each problem size.

CHAPTER 5

Application of (QPO): The Portfolio Rebalancing Problem

This chapter develops a financial application of (QPO), the portfolio rebalancing problem in the presence of transaction costs. Portfolio theory is itself an active area of research within financial optimization [42]. It is also an area where theoretical results have been put into commercial use by market participants. Along these lines, an interesting case study of an institutional MIP approach for the management of large investment portfolios has been published [8]. Before presenting several sets of computational results, we give a brief overview of the idea of efficiency and then formulate the problem of interest as an instance of (QPO) with (DOP) orthogonality.

5.1 The Standard Portfolio Problem and Extensions

Essentially the portfolio optimization problem is to identify the optimal allocation of limited resources among a limited set of investments. Optimality is measured using a tradeoff between perceived risk and expected return. Expected future returns can be based on historical data. Risk can be measured by the variance of those historical returns.

When more than one investment is involved, the covariance among individual investments becomes important. In fact, any deviation from perfect positive correlation allows a beneficial diversified portfolio to be constructed. Efficient portfolios are allocations that achieve the highest possible return for a given level of risk. Alternatively, efficient portfolios can be said to minimize the risk for a given level of return. These ideas earned their inventor a Nobel Prize and have gained such wide acceptance that denumerably-many references could be cited; however, the original source is Markowitz's classic text [31].

One standard formulation of the portfolio problem is:

$$\begin{aligned}
\min_x \quad & \frac{1}{2}x^T V x \\
\text{s.t.} \quad & \mu^T x \geq E_0 \\
& \sum_{i=1}^n x_i = 1 \\
& 0 \leq x_i \leq 1 \quad \forall i
\end{aligned}$$

here n securities are under consideration and E_0 is the required minimum portfolio expected return.

This formulation has a quadratic objective with a set of linear constraints specifying the minimum return for the portfolio and enforcing full investment. The choice to impose zero lower bounds enforces no short-selling. Here μ is the column vector of expected returns and V is the positive semidefinite covariance matrix. The variables x_i are the proportional weights of the i^{th} security in the portfolio.

What we consider is an extension of the basic portfolio optimization problem in which transaction costs are incurred to rebalance a portfolio. That is, transactions are made to change an already existing portfolio, \bar{x} , into a new and efficient portfolio, x . Many market strategists base their investment decisions at least in part on the historical performance of securities. However, historical data becomes less relevant and less useful the more dated it becomes. Therefore, a rebalancing decision of this type would be periodically necessary simply as updated risk and return information is generated with the passage of time. Any alteration to the set of investment choices would also necessitate a rebalancing decision of this type.

5.2 Inclusion of Transaction Costs

The inclusion of transaction costs is an essential element in any realistic portfolio optimization. However, it is important to note that many transaction cost models are possible. Each modeling choice involves a tradeoff between realism and analytic tractability.

Some have modeled transaction costs with integer variables and then implicitly attempted to minimize transaction costs by simply constraining the number of securities allowed in the portfolio [11]. This approach gives rise to quadratic mixed-

integer programming problems; techniques for solving these problems have been developed [29]. Others have moved outside the Markowitz quadratic mean-variance framework and looked at general transaction cost functions possessing concavity or other properties [26].

Staying within the mean-variance framework, our transaction cost model imposes charges as a percentage of the value transacted. Since the weights within the initial portfolio \bar{x} are fixed, transaction costs must be paid only on the movements away from those initial weights. Given that the weights are continuous variables, then the costs we consider may vary continuously as well. In terms of realism, percentages can be specified for each security and can differ between the buying and selling decision for each security.

In addition to the obvious cost of brokerage fees/commissions, here are two examples of other transaction costs that can be modeled in this way:

1. Capital gains taxes are a security-specific selling cost that can be a major consideration for the rebalancing of a portfolio.
2. Another possibility would be to incorporate an investor's confidence in the risk/return forecast as a subjective "cost". Placing high buying and selling costs on a security would favor maintaining the current allocation \bar{x} . Placing a high selling cost and low buying cost could be used to express optimism that a security may outperform its forecast.

5.2.1 Motivation for Percentage Transaction Cost Model

One way to motivate our transaction cost model is to consider transaction costs as an additional security with special risk-return characteristics. The mean expected return of this fictitious security is zero as all funds invested there are certainly lost. The certainty of this loss implies that the variance of this transaction cost security must be zero as well. A further mathematical consequence is that the transaction cost security will be uncorrelated with every other security. It is possible to include such a security directly and then force costs to be paid using one additional constraint.

As before let μ be the n -dimensional vector of returns for the regular securities and V be their covariance matrix. Then augmenting them with transaction costs as the $(n + 1)^{st}$ security results in the following objects, $\tilde{\mu}$ and \tilde{V} :

$$\tilde{\mu} = \begin{bmatrix} \mu \\ 0 \end{bmatrix} \text{ and } \tilde{V} = \begin{bmatrix} V & 0 \\ 0 & 0 \end{bmatrix}$$

Consider the standard portfolio problem for this set of $(n + 1)$ securities:

$$\begin{aligned} \min_x \quad & \frac{1}{2}x^T \tilde{V}x \\ \text{s.t.} \quad & \tilde{\mu}^T x \geq E_0 \\ & \sum_{i=1}^{n+1} x_i = 1 \\ & 0 \leq x_i \leq 1 \quad \forall i \end{aligned}$$

again E_0 is a required minimum portfolio expected return.

How does this help? As written, the transaction cost security would never be active; it offers no return and has no effect on the objective value. Our transaction cost model is incorporated into this problem through the following constraint:

$$x_{n+1} = \sum_{i=1}^n (c_i^{BS}) |x_i - \bar{x}_i|$$

Here c_i^{BS} denotes the opportunity to specify buy and sell transaction costs for each security. This constraint forces investment of the appropriate fraction of portfolio weight changes from the other securities in the transaction cost security. Eliminating this x_{n+1} equality constraint and optimizing w.r.t. the nontrivial covariance matrix leads to the following formulation.

5.2.2 Rebalancing a Portfolio under Percentage Transaction Costs

With this percentage transaction cost model, our problem is expressible as:

$$\begin{aligned} \min_x \quad & \frac{1}{2}x^T Vx \\ \text{s.t.} \quad & \mu^T x \geq E_0 \\ & \sum_{i=1}^n x_i = 1 - \sum_{i=1}^n (c_i^{BS}) |x_i - \bar{x}_i| \\ & 0 \leq x_i \leq 1 \quad \forall i \end{aligned}$$

However, as written there are now piecewise-linear absolute values in the constraints. The following well-known transformation can be used to remove that difficulty by re-expressing the problem in new variables. Let $|x_i - \bar{x}_i| = u_i + v_i$ and $x_i - \bar{x}_i = u_i - v_i$ where $u_i, v_i \geq 0, \forall i$. In terms of these new variables, x_{n+1} the “transaction cost” security obeys the following constraint:

$$x_{n+1} = \sum_{i=1}^n c_i^B u_i + \sum_{i=1}^n c_i^S v_i$$

Here the variables u_i and v_i have clear interpretations as the amount of the i^{th} security bought and sold, respectively.

The transformed problem, written in matrix form, is:

$$\begin{aligned} \min_{u,v} \quad & \begin{bmatrix} \bar{x}^T V & : & -\bar{x}^T V \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \frac{1}{2} \begin{bmatrix} u^T & v^T \end{bmatrix} \begin{bmatrix} V & -V \\ -V & V \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ \text{s.t.} \quad & \begin{bmatrix} -\mu^T & \mu^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \leq \widetilde{E}_0 \\ & \begin{bmatrix} (e_n + c^B)^T & (c^S - e_n)^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = 0 \\ & 0 \leq u_i \leq 1 - \bar{x}_i \\ & 0 \leq v_i \leq \bar{x}_i \quad \forall i \end{aligned}$$

here $\widetilde{E}_0 = \mu^T \bar{x} - E_0$ and e_n is an appropriately sized column vector of ones. Note that the direction of the lone inequality has been reversed to agree with (QPO) standard form.

By varying the return parameter \widetilde{E}_0 and solving multiple instances of this problem, the set of efficient portfolios can be generated. This set, visualized in a risk/return plot, is called the efficient frontier. An investor may decide where along the efficient frontier (s)he finds an acceptable balance between risk and reward.

One general point to be made is that the risk/return plot of an efficient frontier is really just a convenient window on the underlying portfolio weights. For n securities, the portfolio weight vector x exists in \mathbb{R}^n . Further, the “ u, v -space” of buying and selling decision variables for this problem lies in \mathbb{R}^{2n} . So for realistic

numbers of securities, the risk/return plot is a necessity. However, it is important to emphasize that the actual investment decision makes use of the portfolio weights. In other words, the amounts to be bought and sold must be known. For that reason, the optimal solution vector of portfolio weights is required.

5.3 Effect of Changing Parameters

There are three situational parameters that can be studied: the buying (c^B) and selling (c^S) transaction cost percentage vectors as well as the initial portfolio \bar{x} . The effect of each parameter can be studied in isolation. For example, given a constant initial portfolio, the effect of different transaction cost percentages can be explored. Figures 5.1 & 5.4 presents such a scenario where the initial portfolio was equally divided amongst all securities and all costs are uniform (ie. a simple scalar multiple $c^B = c^S = c \cdot e_n$). For reference purposes, all the efficient frontiers plotted in Sections 5.3 - 5.6 involve a set of nine securities and use annualized return data from page 13 of Markowitz's text [31]. Additional information about the Markowitz dataset can be found in Appendix A.

As will be discussed, these curves shown in Figure 5.1 are “efficient” only in the mean-variance sense. They are not efficient once nonzero transaction costs are involved. It is the goal of the next section to illustrate how (DOP) orthogonality constraints are essential for identifying the frontier of portfolios which are efficient both in terms of mean-variance and in terms of the transaction costs paid to rebalance.

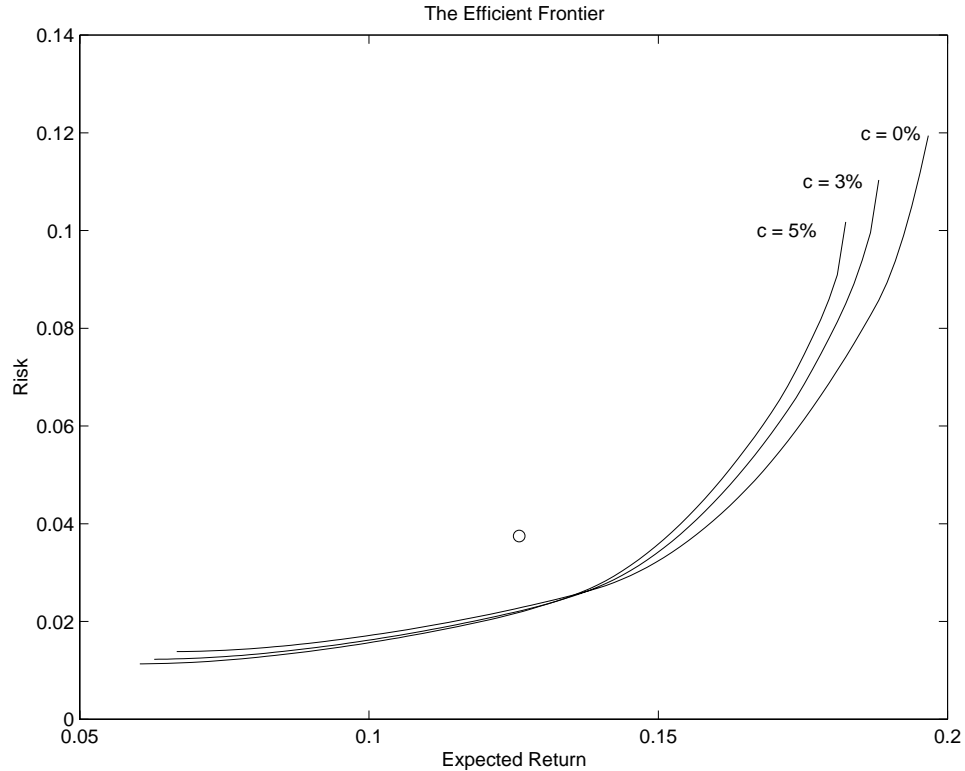


Figure 5.1: The initial portfolio is located by a circle. Notice that as the level of transaction costs c increases, the frontier shifts left corresponding to a decrease in return for a given level of risk. Also, notice that the efficient frontiers for different transaction costs cross.

5.4 Crossover Among “Efficient” Frontiers

From a conceptual point of view, the crossover observable in Figure 5.1 was troubling. We determined that two issues were involved. The first issue leads directly to the necessity of (DOP) orthogonality constraints. The second issue involves the appropriate scaling of the portfolio risk measurement and is discussed in Section 5.4.3.

5.4.1 Necessity of (DOP) Orthogonality

We had imposed independent range restrictions on the buying and selling decision variables in the previous formulation of Section 5.2.2. However, this set of constraints is not sufficient. Figure 5.1 demonstrates that simultaneous buying and

selling within a single security was sometimes found to be optimal. Such a condition was preferred because decreases in the quadratic objective were sometimes achieved through the cross-terms involving both u and v .

To elaborate on this remark, the transformation introduced in Section 5.2.2 can be re-expressed as follows: $x_i = \bar{x}_i + u_i - v_i$. In words, this says that the final position in the i^{th} security is equal to the original position plus whatever is bought less whatever is sold. In this expression, notice that mathematically an infinite number of $\{u_i, v_i\}$ combinations are possible for each fixed value of x_i . Any combination of u_i and v_i with simultaneous nonzeros corresponds to a decision which simultaneously buys and sells within a single security to arrive at its final allocation.

However, it cannot be “efficient” to incur costs both buying and selling the same security at the same time. Intuitively, only one decision will be made. To rebalance: the decision would be to either buy or sell and never do both. This means that an orthogonality relationship must be maintained between the buying and selling decisions for a given security. The necessary constraints have the form $u_i v_i = 0 \forall i$. Figure 5.2 illustrates the graph structure of these constraints.

Making this change, we arrive at the following restatement of the portfolio rebalancing problem:

$$\begin{aligned}
& \min_{u,v} \left[\bar{x}^T V \quad : \quad -\bar{x}^T V \right] \begin{bmatrix} u \\ v \end{bmatrix} + \frac{1}{2} \begin{bmatrix} u^T & v^T \end{bmatrix} \begin{bmatrix} V & -V \\ -V & V \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\
& \text{s.t.} \quad \begin{bmatrix} -\mu^T & \mu^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \leq \widetilde{E}_0 \\
& \quad \begin{bmatrix} (e_n + c^B)^T & (c^S - e_n)^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = 0 \\
& \quad 0 \leq u_i \leq 1 - \bar{x}_i \\
& \quad 0 \leq v_i \leq \bar{x}_i \\
& \quad u_i v_i = 0 \quad \forall i
\end{aligned}$$

Notice that with the inclusion of orthogonality constraints, this becomes an instance of (QPO). In fact it was this specific formulation that originally motivated the research contained in this thesis.

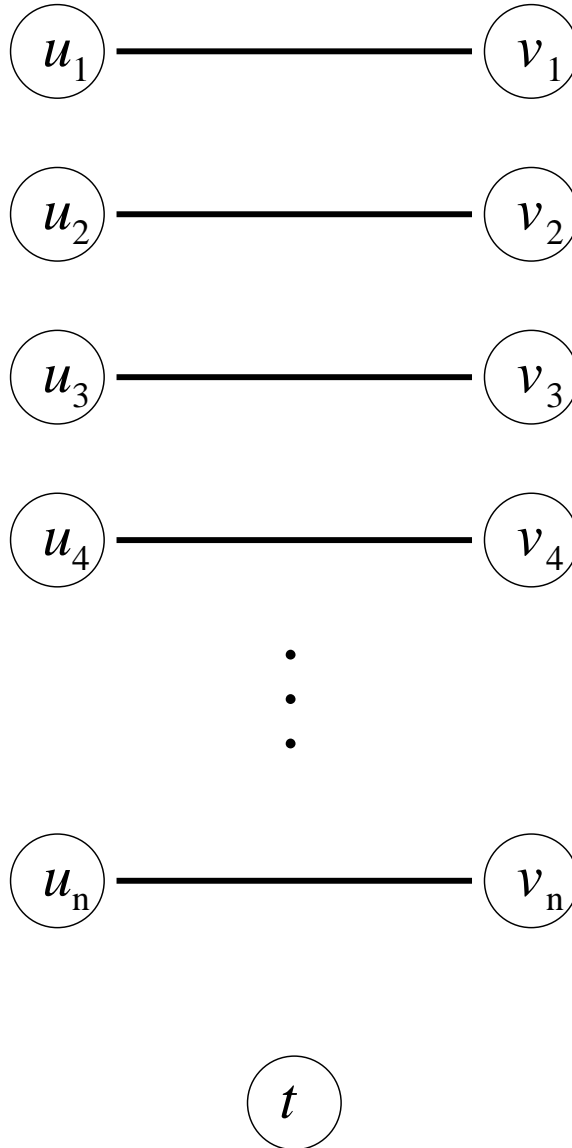


Figure 5.2: This graph exemplifies the (DOP) structure of the orthogonality constraints required for this application. For n securities there are $2n$ perfectly matched variables in the graph. The variable t is a scaling variable that is discussed in Appendix C; it is the only isolated node.

5.4.2 Nonconvexity of the Feasible Region

One general property of (QPO) that can be clearly demonstrated here in the context of the portfolio rebalancing problem is that the feasible region is nonconvex. This nonconvexity is what makes (QPO) an NP-hard optimization problem. For the portfolio rebalancing application, the nonconvexity results from imposing the (DOP) orthogonality constraints in the presence of nonzero transaction costs.

Figure 5.3 shows a simple two dimensional example plotted in the space of portfolio weights. For presentation purposes, the minimum portfolio return constraint is not shown. The initial investment is equal between the securities and costs are kept uniform. Analytically, we let $\bar{x} = [.5, .5]^T$ and $c^B = c^S = c \cdot [1, 1]^T$. The feasible region is observed to “break” at \bar{x} . As you can also see, each of the faces correspond to an orthogonal set of movements in the space of buying and selling decision variables. Using the terminology of this thesis, in this two dimensional example there are two possible active sets that collectively exhaust all the feasible buying/selling combinations. We will return to the geometry of the feasible region in order to motivate our response to degenerate decisions in Chapter 6.

5.4.3 Issue of the Remaining Crossover: Fractional Objective

Enforcing orthogonality fixed a modeling error and led us to study (QPO) but did not preclude the frontiers from crossing over as transaction costs increased. Looking for a financial interpretation of the crossover led us to the following conclusion regarding the quantity that should actually be optimized. To this point, we have been optimizing the standard risk measure for efficient frontiers, that is:

$$\frac{1}{2}x^T V x$$

When there are no transaction costs to be paid, one dollar is always available for investment, i.e. $(\sum_{i=1}^n x_i = 1)$. This assumption is implicit in the standard risk measure. However, for nonzero transaction costs that implicit assumption is no longer valid. One dollar is not available for investment, costs will be paid to rebalance. So, what quantity should be optimized?

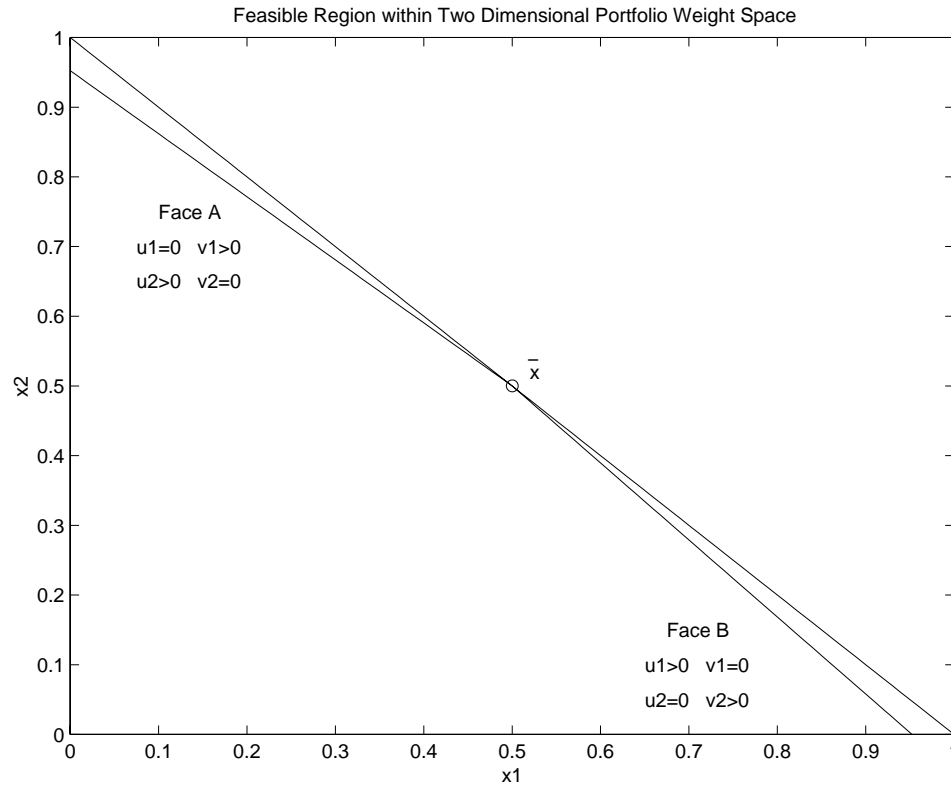


Figure 5.3: In this figure, the $c = 0\%$ feasible region is the above line segment from $x_1 + x_2 = 1$ which is a convex set. The feasible region when $c = 5\%$ consists of the two labeled faces that intersect at \bar{x} . This is no longer a convex set.

Our conclusion, the appropriate objective is:

$$\frac{\frac{1}{2}x^T V x}{\{1 - x_{n+1}\}^2}$$

Here x_{n+1} is again the amount allocated to the transaction cost security. Therefore $\{1 - x_{n+1}\}$ is the actual amount available for investment. So we are choosing to scale the standard risk measurement by the square of the dollar amount actually invested.

Analytically, notice that with zero transaction costs then $x_{n+1} \equiv 0$ and we recover the standard risk measurement. So our choice does pass the first test required of any theoretical extension; recover the previous result. This choice also makes dimensional sense given the quadratic numerator.

Our choice of this fractional objective function also makes intuitive sense. For nonzero transaction costs, there are conflicting effects at work within the portfolio. Thinking of fixed \bar{x} then as the transaction cost percentage is increased, you expect smaller absolute amounts of principal will be available for investment. But in order to get the same payoff ($\mu^T x$) on a smaller amount of principal you will need to reach for higher returns. This should correlate to taking on higher levels of risk. Our fractional choice effectively boosts the risk measurement for these transaction cost depleted portfolios.

Optimizing this fractional objective creates the situation where the $c = 0\%$ frontier extends furthest into the risk/return plane. Other transaction cost efficient frontiers, abbreviated TCEF, are pulled back from that limit as seen in Figure 5.4. That the frontiers no longer cross leads us to the interpretation that transaction costs reduce the range of investment choice.

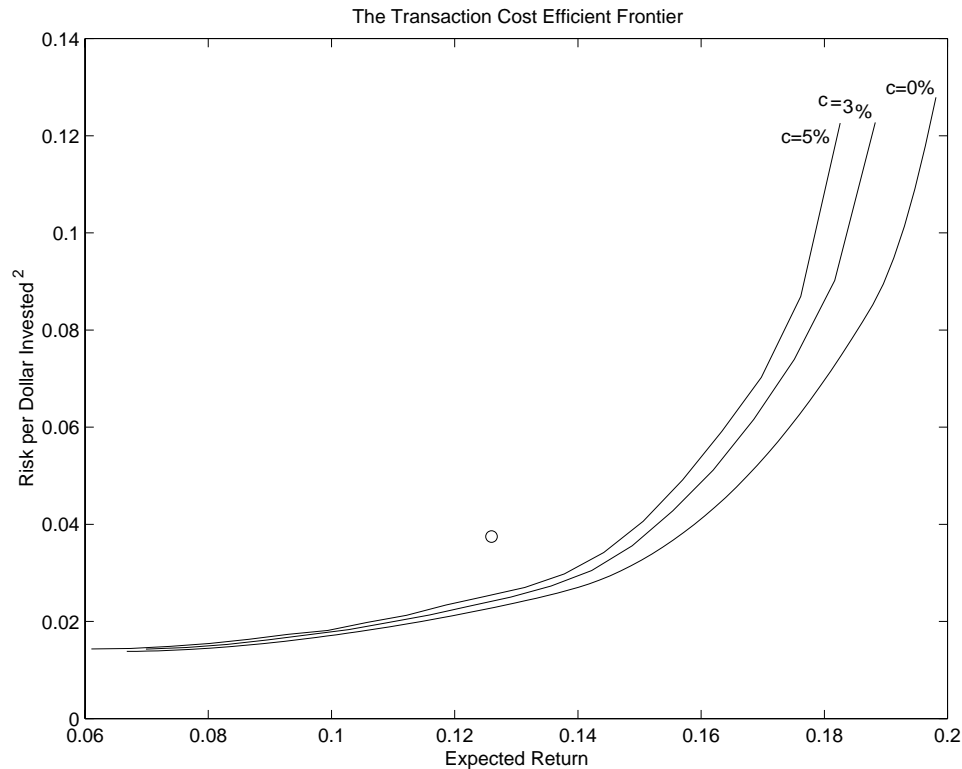


Figure 5.4: The initial portfolio is again located by a circle. Notice that as the level of transaction costs c increases, the curves shift left. However, crossover no longer occurs. Increased transaction costs reduce investment choice.

5.5 Performance Analysis for the Markowitz $c=5\%$ TCEF

Figure 5.4 on the previous page presents computational results illustrating the effect of transaction costs on the transaction cost efficient frontier. That information is of financial interest; another type of information is interesting mathematically. Each point on a nonzero transaction cost frontier corresponds to a separate instance of (QPO). Each instance is a separate test of how well our solution strategy performs. Refreshing the reader, our strategy is to use the semidefinite relaxation (rSDP) as the basis for selecting a promising set of QP-subproblems. The best observed result subject to degeneracy is returned as optimal. So looking at a specific frontier or even just at a specific point, performance information regarding our solution strategy can be presented.

Figure 5.5 on the following page shows the same $c=5\%$ TCEF as Figure 5.4. However in Figure 5.5, the mathematically interesting performance information is presented. For this TCEF, the Spectral Bundle solver [23] was allowed a 10 minute time limit to solve each instance of (rSDP). For all instances, (rSDP) was solved to optimality before that time limit was reached. Again a conservative set of tolerances ($L_{tol} = 10\%$ and $U_{tol} = 90\%$) was used to make activity decisions. We take this opportunity to prepare the reader by explaining the legend used in Figure 5.5:

- “Hmin(QPsubs)” points are found using our solution strategy.
- “min(QPSubs)” points result from exhaustively considering all possibilities.
- Finally, the series “rSDP Optimal” locates the optimal value of the relaxation.

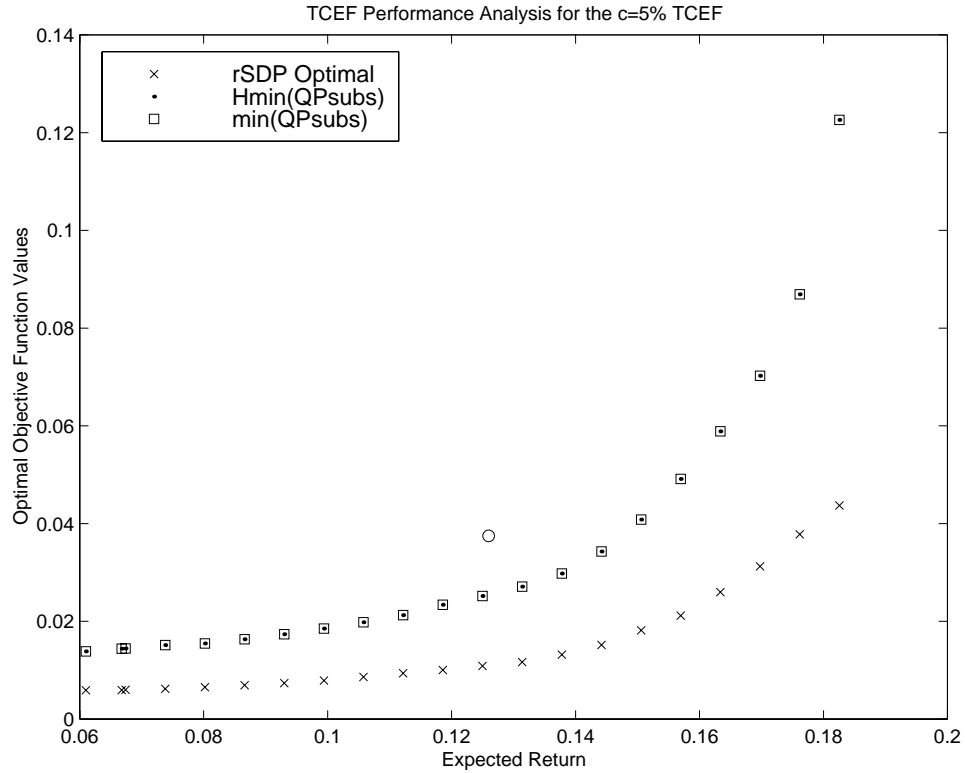


Figure 5.5: The circle again locates the initial portfolio. This figure presents information regarding the performance of our solution strategy.

Looking at Figure 5.5, several observations can be made. To provide orientation, comparisons can be made for each level of required expected return. This corresponds to vertical comparisons in the plot. First, notice that our solution strategy correctly identified the true optimal solution in every instance along this frontier. This is indicated by the “dot within a square.” A second observation is that the optimal value of (rSDP) does appear to be a consistently tight lower bound on the true optimum.

Table 5.1 on the following page contains the numerical results shown above. It also quantifies the assertion that our strategy considered a relatively small number of QP-subproblems in order to find the optimal solution. To explain the lone “n.a.” entry in Table 5.1, the highest returning point on any TCEF is directly solvable as a linear program. This topic is addressed in Appendix B. Therefore, equivalent performance information regarding the number of QP-subproblems is not applicable.

Required Expected Return (%)	(rSDP) Objective Value	Minimal QP-subproblem Objective Value	Number of QP-subproblems Considered
18.26	0.043725	0.122595	n.a.
17.62	0.037793	0.086903	16
16.98	0.031234	0.070248	16
16.34	0.025977	0.058898	16
15.70	0.021162	0.049131	16
15.06	0.018161	0.040815	16
14.42	0.015173	0.034284	16
13.78	0.013183	0.029781	16
13.14	0.011649	0.027103	16
12.50	0.010877	0.024602	16
11.86	0.010024	0.023381	32
11.22	0.009383	0.021267	32
10.58	0.008607	0.019820	32
9.94	0.007897	0.018517	32
9.30	0.007363	0.017363	32
8.66	0.006915	0.016326	32
8.02	0.006528	0.015471	16
7.38	0.006182	0.015153	16
6.74	0.005994	0.014462	16
6.68	0.005937	0.014426	16
6.10	0.005909	0.013850	8

Table 5.1: Since 9 securities were involved, the fourth column compares against a total of $2^9 = 512$ possible QP-subproblems. Notice that for most cases, correct decisions were made for 5 out of the 9 securities. No degenerate solutions were encountered during this calculation.

5.5.1 Examination of the QP-subproblems Selected by the Strategy

Table 5.1 demonstrates that our strategy selected a relatively small subset from the set of all possible QP-subproblems. However, the optimal subproblem was located within that subset. Was this simply luck or can we provide some additional support for the performance of our strategy? Support can and does come from an analysis of the distribution of the QP-subproblems selected by our strategy against the set of all possible subproblems.

We now choose to focus on a single representative point rather than the entire Markowitz $c=5\%$ TCEF. The point chosen corresponds to a required expected return of 12.50%. This is close to the return already offered by the initial portfolio since $\mu^T \bar{x} = 12.60\%$. Exhaustively searching all 512 possible QP-subproblems identifies that only 290 subproblems are feasible. Each feasible QP-subproblem has an optimal objective value which we sort to identify the 10 Best QP-subproblems. Here, “Best” refers to the property that from the set of all possibilities, these QP-subproblems have the 10 smallest optimal objective values.

A similar examination of the 16 QP-subproblems that were selected by our strategy finds that 9 are feasible. The optimal objective values of these 9 QP-subproblems can also be sorted by objective value. Table 5.2 presents a side-by-side comparison of the results.

Optimal Objective Values for the 10 Best QP-subproblems	Optimal Objective Values for QP-subproblems Selected by the Strategy	Comment on Relationship
0.024602	0.024602	Strategy selected true optimal solution.
0.024703		Not selected, see Figure 5.6.
0.024722	0.024722	Selected by our strategy.
0.024737	0.024737	Selected by our strategy.
0.024875	0.024875	Selected by our strategy.
0.024877	0.024877	Selected by our strategy.
0.025168	0.025168	Selected by our strategy.
0.025186		
0.025238		
0.025240		
	0.033665	Feasible but not close to optimal.
	0.034617	Feasible but not close to optimal.
	0.035161	Feasible but not close to optimal.

Table 5.2: Notice that the QP-subproblems selected by our strategy are densely clustered near the true optimum. More specifically, the strategy selected the true optimal and 6 other subproblems located within 2% of the true optimum.

In Table 5.2, the “2nd Best” QP-subproblem with objective value 0.024703 was not selected by our strategy. This would certainly seem to weaken this analysis. However, looking at the actual portfolios which are optimal in each subproblem is enlightening. Figure 5.6 below presents that information. The vertical ordering and side-by-side comparison format of Table 5.2 is replicated in Figure 5.6. Symbols indicate whether the optimal allocation of a particular security was unchanged, increased, or decreased from its initial value. The symbols used are: bought \triangle , sold ∇ , unchanged \square . The absence of a symbol means that the entire allocation of that particular security was liquidated.

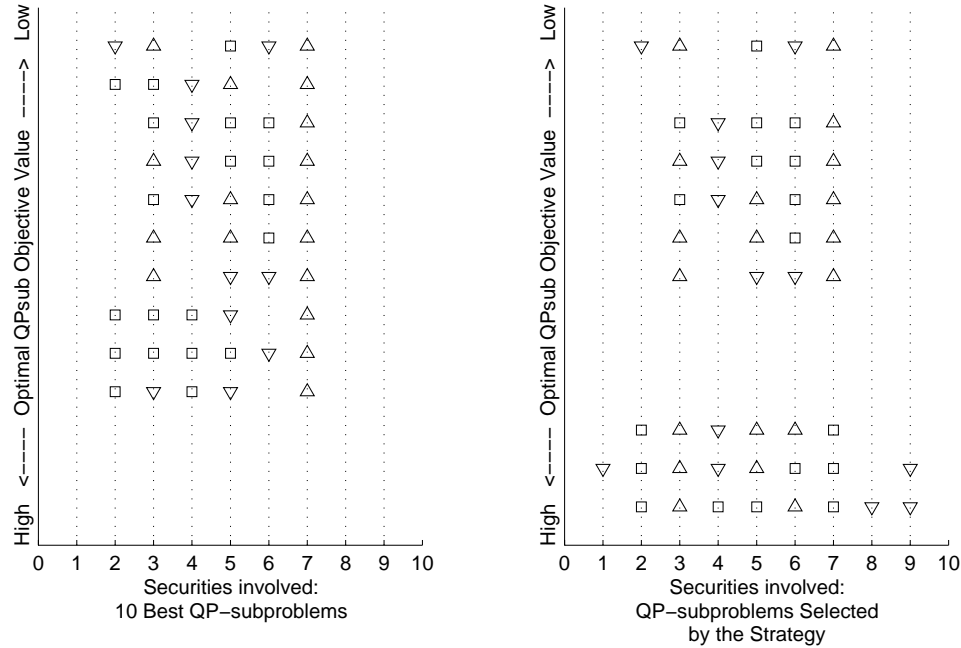


Figure 5.6: Securities numbered 1 through 9 exist. No activity decisions were made for securities #3, 5, 6, and 7. The first and second rows of the lefthand panel are of particular relevance.

Notice that the optimal solution of the 2nd Best QP-subproblem is feasible in the true optimum. Table 5.3 highlights this by showing the active set choices for these subproblems. This subproblem was not selected because it violated the activity decision to sell security #2. As evidenced by the true optimum, that sell decision was in fact correct. So, an analysis of the optimal solutions supports the performance of our strategy.

Objective Values for the 2 Best QP-subproblems	Security Number								
	1	2	3	4	5	6	7	8	9
0.024602	Sell	Sell	Buy	Sell	Buy	Sell	Buy	Sell	Sell
0.024703	Sell	Buy	Buy	Sell	Buy	Sell	Buy	Sell	Sell

Table 5.3: This table contains the active set choices for the 2 Best QP-subproblems. Notice that the slightly suboptimal QP-subproblem solution is feasible in the true optimal. All decisions are the same except for security #2. Since this subproblem was constrained to buy security #2, the slightly suboptimal solution is a “Buy=0” decision.

Stated in another way, the slightly suboptimal 2nd Best QP-subproblem was not selected because it lies on the incorrect side of the $x_2 = \bar{x}_2$ decision boundary. Across that boundary, on a neighboring face, lies the true optimum. These geometric ideas are made more rigorous in Chapter 6. However to conclude this analysis, we observe that our strategy selected a subset of QP-subproblems which disproportionately populates the most promising region of the search space. This is exactly the behavior which was the original goal of our solution strategy.

5.5.2 Presentation of the Markowitz $c=5\%$ TCEF Portfolios

As was mentioned earlier, the actual real-world rebalancing process requires information about the actual amounts of each security that should be bought or sold. Before leaving this TCEF, Figure 5.7 on the following page shows which securities are involved at each point along the TCEF. We use the same symbology that was introduced for Figure 5.6. However, the vertical scale is now the minimum required expected return (ie. the horizontal coordinate) of the efficient frontier plots.

Figure 5.7 contains a great deal of useful information which can be accessed by visual inspection. We will not try to exhaust all possibilities but describe a few examples. Looking down a column, changes occurring in the portfolio as the required expected return is decreased can easily be seen. Also, horizontal comparisons between panels allow for comparisons to be made between the no-cost and $c=5\%$ frontiers.

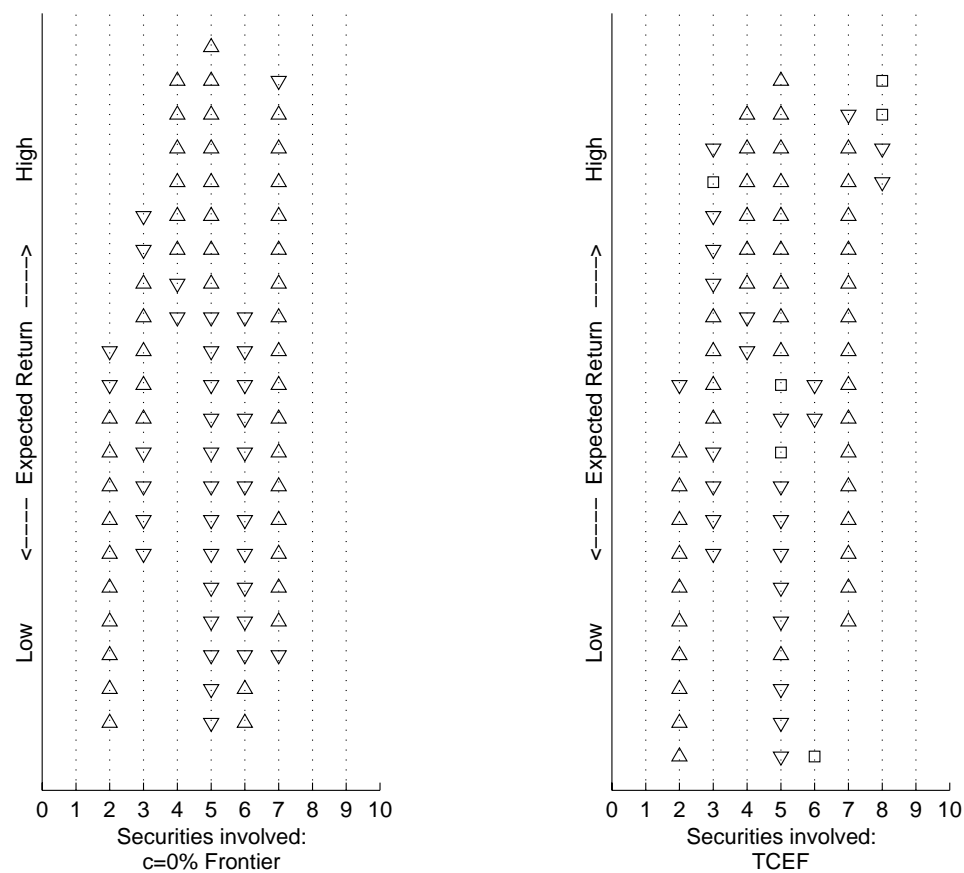


Figure 5.7: These panels show which securities are involved along the no-transaction cost efficient frontier (left) and the $c=5\%$ TCEF (right). The symbology is explained on page 64.

Two observations regarding Figure 5.7 deserve special mention:

- The highest returning portfolio on the no-cost frontier has no counterpart on the $c=5\%$ TCEF. Likewise, the least risky-least returning point on the $c=5\%$ TCEF has no counterpart on the no-cost frontier. This is a manifestation of the leftward shift seen in Figure 5.4.
- With no transaction costs the highest returning portfolio is always full investment in the single highest yielding individually efficient security. This is security #5 for the Markowitz dataset. However with the introduction of nonzero transaction costs, that is no longer the case.

A general observation is that the portfolios along the TCEF are not simply related to the portfolios along the no transaction cost efficient frontier. Sometimes, entirely new securities are involved. Sometimes, buy and sell decisions are reversed. So Figure 5.7 highlights that the introduction of costs changes the portfolio rebalancing problem dramatically and that the optimal solutions are also quite different.

5.6 Performance Analysis for an Additional Markowitz TCEF

The previous performance analysis for the $c=5\%$ Markowitz TCEF was varied and highly detailed. For the sake of the reader's attention span as well as the author's patience, we will not continue at that level of detail. However, we will present additional computational results which highlight other issues. For example, the results of this section are empirical evidence why our solution strategy must address the issue of degeneracy. Our discussion of the geometric interpretation and algorithmic impact of degeneracy can be found in Chapter 6.

The uniform portfolio allocation and cost structure of Section 5.5 is not the most realistic test for our solution strategy. Real portfolios change their relative composition over time as the better-returning securities grow more than those performing less well or those securities that decline outright. So nonuniform initial allocations are worth consideration. Also, to take full advantage of the flexibility of our cost model, nonuniform costs will be considered as well.

The TCEF presented in this section is a second example that makes use of the Markowitz dataset. This means that exhaustive confirmation is again available. All nine securities are involved initially. The allocations were randomly selected as were the buying and selling costs. Those costs were kept on the order of a few percent. Table 5.4 below contains the values of these parameters.

Security Number	Initial Allocation (%)	Buying Cost (%)	Selling Cost (%)
1	13.33	4.75	2.22
2	29.03	1.16	3.08
3	1.88	3.03	3.96
4	11.46	2.43	4.61
5	26.42	4.46	3.69
6	0.32	3.81	0.88
7	4.51	2.28	2.03
8	6.59	0.09	4.68
9	6.46	4.11	4.58

Table 5.4: The initial allocations were randomly selected. Transaction costs were also randomly selected but kept below 5%.

As an interesting aside, note that the inclusion of entirely new securities in a portfolio can be accommodated within the rebalancing decision. Those entirely new securities are already present in the existing portfolio; they simply have zero weight. This adds a new dimension to the rebalancing decision. By extending this reasoning, one can consider the issue of whether or not to rebalance. We will return to this issue briefly in Chapter 7.

Figure 5.8 below presents performance information about our strategy for this additional set of parameters. Again, the Spectral Bundle solver [23] solved all (rSDP) instances to optimality before a time limit of 10 minutes was reached. Conservative tolerances ($L_{tol} = 10\%$ and $U_{tol} = 90\%$) were used to make activity decisions. Notice that (rSDP) is again a consistently tight lower bound and that the optimal solution is successfully identified. Table 5.5 on the following page contains the numerical results.

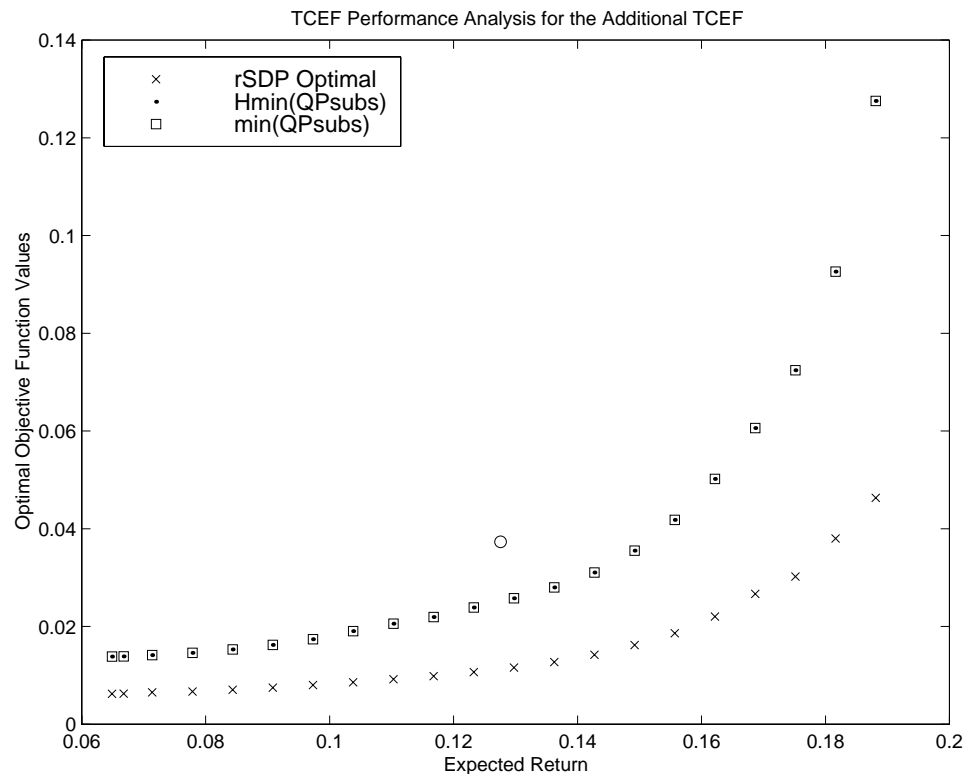


Figure 5.8: Whenever degeneracy was detected additional QP-subproblems were generated. The circle again locates the initial portfolio.

Required Expected Return (%)	(rSDP) Objective Value	Minimal QP-subproblem Objective Value	Total Number of QP-subproblems Considered	
			Conservative Approach	Aggressive Approach
18.81	0.046333	0.127526	n.a.	11
18.16	0.038014	0.092625	32	
17.51	0.030219	0.072423	16	
16.87	0.026677	0.060602	8	
16.22	0.022051	0.050215	16	
15.57	0.018635	0.041821	32	
14.92	0.016205	0.035531	16	
14.27	0.014231	0.031043	32	
13.62	0.012710	0.027989	16	
12.97	0.011593	0.025796	16	
12.33	0.010660	0.023906	16	
11.68	0.009846	0.021945	64	
11.03	0.009209	0.020593	32	
10.38	0.008581	0.019065	32	
9.73	0.008011	0.017382	32	
9.08	0.007492	0.016233	32	
8.43	0.007068	0.015317	16	
7.79	0.006689	0.014609	16	
7.14	0.006532	0.014148	16	
6.68	0.006279	0.013899	16	
6.49	0.006243	0.013867	8	

Table 5.5: Both responses to degeneracy are tabulated. Since these two approaches differ only for degenerate points, the number of QP-subproblems actually considered is only stated in the fourth column for nondegenerate points. For the point where degeneracy was encountered, separate totals are provided for each approach. Notice that the aggressive approach solves fewer QP-subproblems but still identified the true optimum in this case.

This TCEF was chosen for presentation simply to illustrate the need for a respond to degeneracy. The conservative and aggressive approaches referred to in this section are explained in detail in Chapter 6. To conclude this section, Figure 5.9 presents the portfolios on this additional Markowitz TCEF. As before, visual inspection reveals that they are quite different from the no transaction cost efficient frontier.

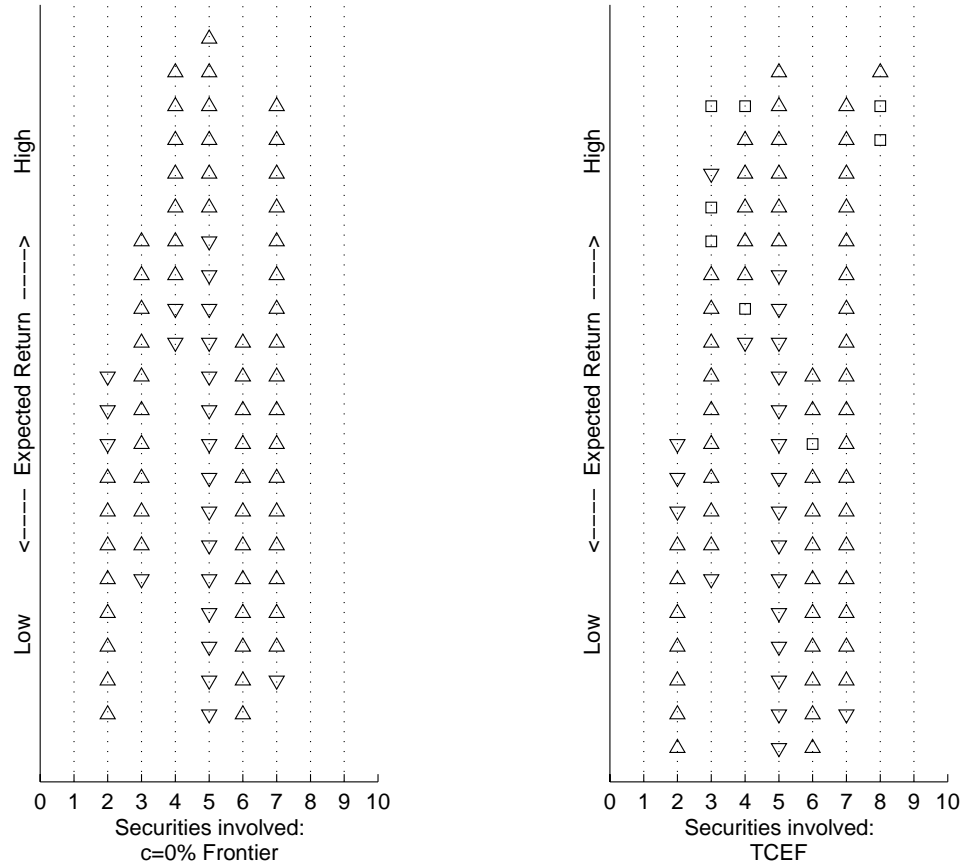


Figure 5.9: These panels show which securities are involved along the no-transaction cost efficient frontier (left) and the additional TCEF (right). The symbology is explained on page 64 and the parameter values are given in Table 5.4.

5.7 Rebalancing a Dow Jones Portfolio

The Markowitz dataset used in the previous sections was small enough that exhaustive confirmation was possible. In the language of Chapter 4, the “naïve approach” for solution has been available for this smaller dataset. That alternate method of solution was critical for research purposes during the development and refinement of our solution strategy. Hopefully, it also helped convince the reader of the validity of our solution strategy. In this final set of computational results, we demonstrate the effectiveness of our strategy on a larger and so more realistic dataset.

We applied our solution strategy to the problem of rebalancing portfolios composed of the 30 stocks which currently make up the Dow Jones Industrials Average. All securities were involved initially. It is important to stress that with a portfolio of 30 securities, there are $2^{30} \sim 10^9$ or over a billion possible QP-subproblems. Obviously, exhaustive confirmation is not available. However, the results we present testify to the continued success of our strategy.

The same order of presentation is used for this Dow Jones TCEF as was used with the previous Markowitz dataset TCEF. On subsequent pages, we present:

- the values used for parameters \bar{x}, c^B, c^S in Table 5.6,
- a plot demonstrating the performance of our strategy in Figure 5.10,
- the underlying numerical results in Table 5.7,
- and visualizations of the actual portfolios selected in Figures 5.11 & 5.12.

For this TCEF, the Spectral Bundle solver was allowed 30 minutes to solve each instance of (rSDP). However, no instance was solved to optimality before that time limit expired. As the caption of Figure 5.10 notes, this degrades how tight of a lower bound (rSDP) provides for some points along the TCEF. However, the use of suboptimal (rSDP) solution matrices does not appear to have degraded the validity or performance of our decision-making process. Large numbers of activity decisions were made and no degenerate points were encountered. The same conservative set of decision tolerances was used.

Security Number	Initial Allocation (%)	Buying Cost (%)	Selling Cost (%)
1	2.74	4.75	0.08
2	4.97	1.16	3.73
3	4.54	3.03	2.23
4	3.56	2.43	4.66
5	4.52	4.46	2.33
6	3.64	3.81	2.09
7	1.89	2.28	4.23
8	1.60	0.09	2.63
9	1.88	4.11	1.01
10	2.95	2.22	3.36
11	4.01	3.08	4.19
12	1.71	3.96	0.10
13	4.63	4.61	3.41
14	3.14	3.69	1.90
15	2.04	0.88	4.16
16	3.88	2.03	2.51
17	3.02	4.68	3.55
18	2.46	4.58	2.14
19	3.83	2.05	1.52
20	3.43	4.47	0.95
21	4.39	0.29	0.97
22	5.28	1.76	3.41
23	2.88	4.07	1.51
24	4.86	0.05	2.71
25	0.95	0.69	0.75
26	5.41	1.01	3.49
27	1.50	0.99	1.89
28	1.39	3.02	4.30
29	4.83	1.36	4.27
30	4.07	0.79	2.97

Table 5.6: Parameters used for the Dow Jones Portfolio

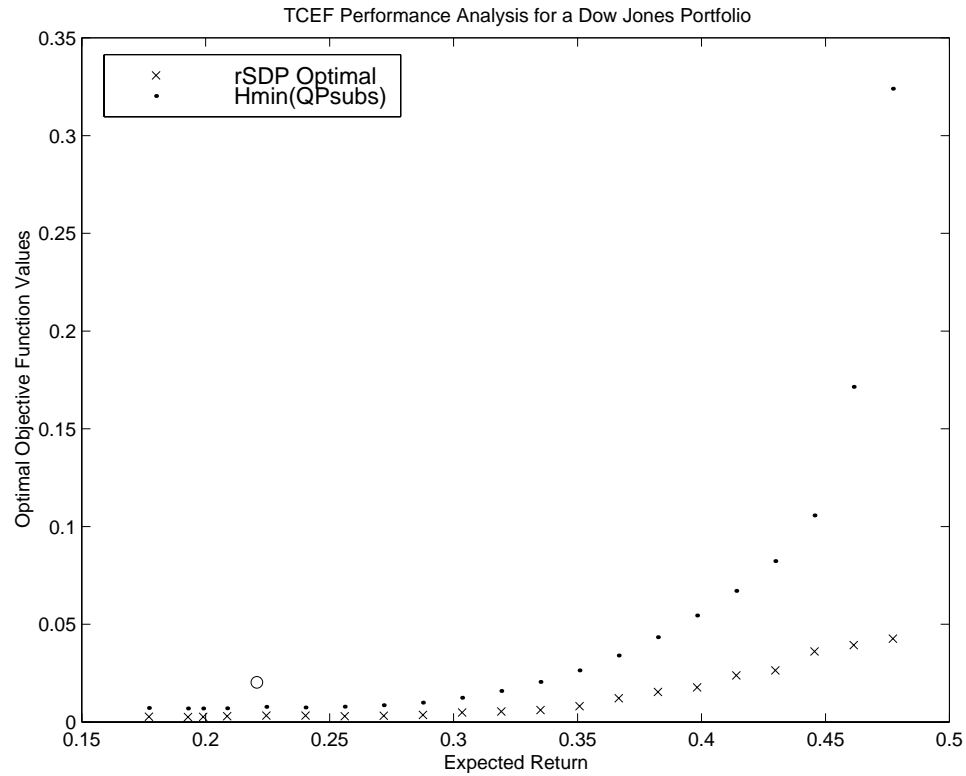


Figure 5.10: The “min(QPsubs)” series cannot be provided for the Dow Jones dataset. The circle again locates the initial portfolio. Notice that the (rSDP) optimal objective value does not mirror the shape of the TCEF, particularly for higher required expected returns. This is a noticeable difference from Figures 5.5 & 5.8. However, the TCEF itself has no visible defects. This provides strong empirical support for our strategy.

Required Expected Return (%)	(rSDP) Objective Value	Minimal QP-subproblem Objective Value	Number of QP-subproblems Selected by the Strategy
47.72	0.042696	0.324000	n.a.
46.14	0.039363	0.171472	256
44.56	0.036101	0.105743	256
42.98	0.026411	0.082315	256
41.40	0.023847	0.067093	256
39.82	0.017756	0.054479	256
38.24	0.015437	0.043378	256
36.66	0.012181	0.034012	512
35.08	0.008156	0.026464	1024
33.50	0.006154	0.020584	2048
31.92	0.005408	0.015929	1024
30.34	0.004849	0.012424	2048
28.77	0.003656	0.009964	256
27.19	0.003308	0.008580	512
25.61	0.003284	0.007910	256
24.03	0.003193	0.007751	2048
22.45	0.003009	0.007494	16384*
20.87	0.002962	0.007234	1024
19.90	0.002736	0.007011	256
19.29	0.002661	0.006952	256
17.71	0.002534	0.006923	256

Table 5.7: For this dataset, the fourth column compares against a total of $2^{30} \sim 10^9$ or over a billion possible QP-subproblems. For most cases, activity decisions were made for 22 out of the 30 securities. The third column corresponds to the “Hmin(QPsubs)” series in Figure 5.10. No degenerate solutions were encountered. For the point marked by the asterisk, the calculation was stopped after 2048 QP-subproblems were considered.

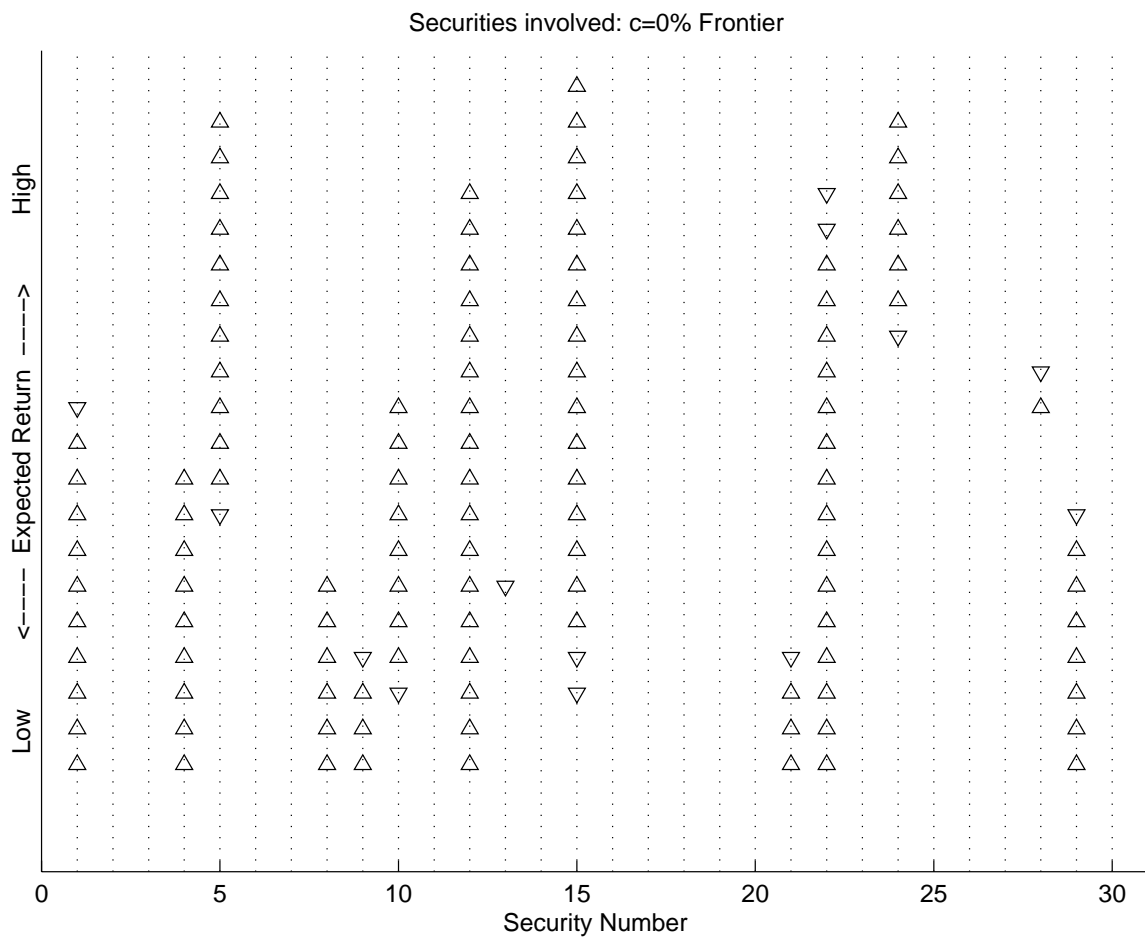


Figure 5.11: Involved Securities for the Dow Jones Portfolio: This figure shows which securities are involved along the $c=0\%$ frontier.

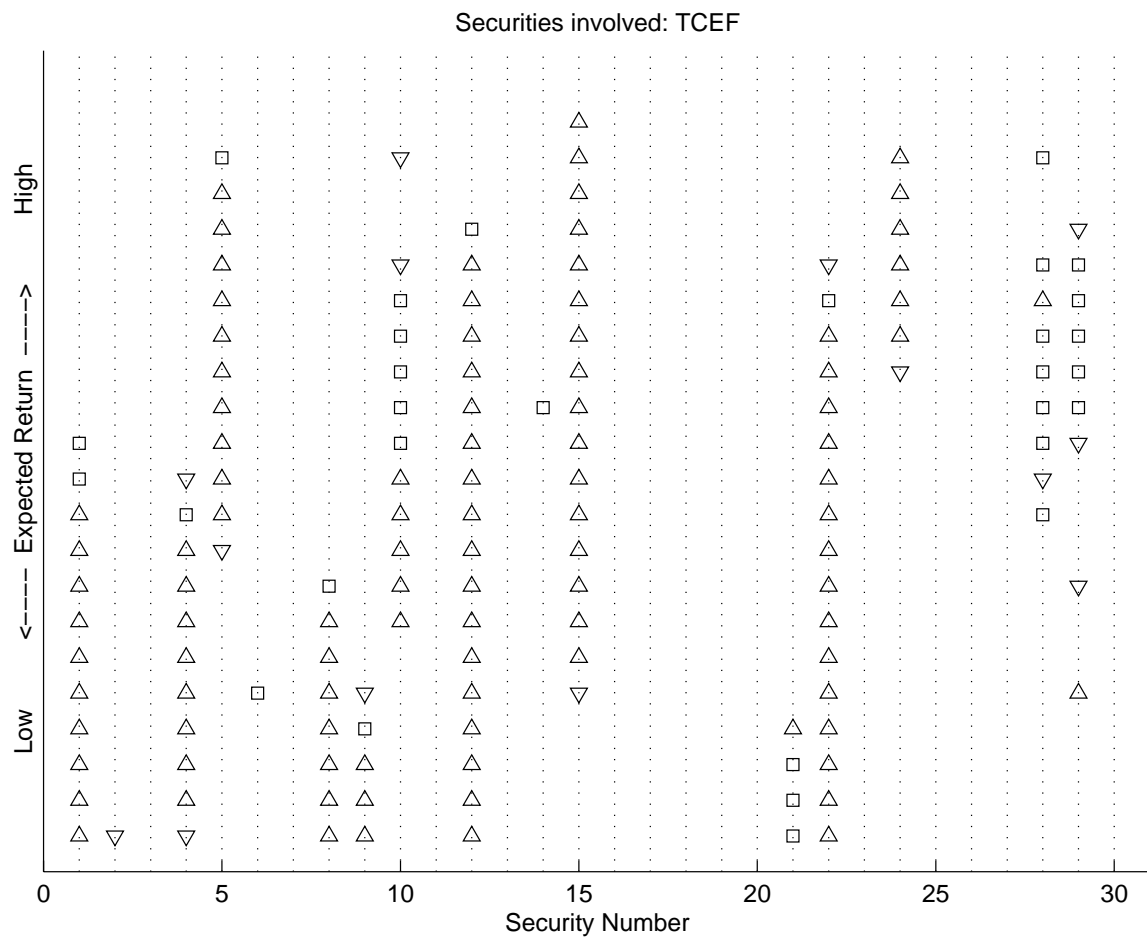


Figure 5.12: Involved Securities for the Dow Jones Portfolio: This figure shows which securities are involved along the TCEF. Here, overlay comparisons with Figure 5.11 are possible.

Organizational Note:

For reference, additional information about the Dow Jones dataset is presented in Appendix A. Following that, two points of tangential interest that were mentioned in this chapter are included as appendices. Appendix B presents the approaches used to identify the endpoints of a TCEF. This is a nontrivial problem in itself and is interesting in its own right. Appendix C presents the transformation that was used to restate the fractional objective function as a simple quadratic expression. This transformation allows us to implement the scaled risk measurement of Section 5.4.3.

CHAPTER 6

The Issue of Degenerate Decisions

To set the stage for this chapter, take the following perspective on the material of the preceeding two chapters. In Chapter 4, our solution strategy and decision-making heuristic were presented. The key attribute there is that our approach can be classified as an enumerative algorithm. In Chapter 5, several sets of computational results were presented and the performance of our solution strategy was analyzed. For the purposes of this chapter, the insight to draw from those computational results, specifically Section 5.6, is simply that degenerate decisions can occur and a response to degeneracy is required to make our strategy more robust.

This chapter focuses on how an understanding of the geometry of the feasible region allows us to sensibly grow the search space when degeneracy is found. This chapter concludes with an analysis of different ways to grow the search space.

6.1 Respond by Growing the Search Space

Our earlier discussion of solution strategy was centered on how the definitive cases are treated. We basically used the definitive cases to partition the search space. Solving (rSDP) can be thought of as a means for identifying the most-promising portion of that search space. The corresponding set of QP-subproblems was found to remain tractable and sometimes -but not always- contained the optimal solution.

We observed that degenerate decisions did occur. That experience allows us to elaborate on our previous definition of a degenerate decision. Refreshing the reader, a degenerate decision is one where the active member of an orthogonal pair is found to be zero in the solution returned as optimal. In the context of our solution strategy, degeneracy has the following effects. At the very least, degeneracy causes us to exclude more possibilities from consideration than can really be justified. At worst, degeneracy can be one obstacle that prevents us from identifying the optimal solution. Picking up on a point from Chapter 4, degeneracy relates to the accuracy -or better said- inaccuracy of our information regarding the optimal active set.

Our response to degenerate decisions will be to enumerate more possibilities. However, we obviously want to generate and solve as small a total number of QP-subproblems as possible. So if we respond to degeneracy by solving additional QP-subproblems, which problems are sensible? It turns out that the geometry of the feasible region gives us a rationale for which additional QP-subproblems should be considered.

6.2 Visualizing the Geometry

In Chapter 4, we established a correspondence between each possible active set and a convex QP-subproblem. Here that correspondence is extended to a third geometrical counterpart, a face within the feasible region. The discussion that follows is in terms of the financial application of Chapter 5; however, the insight holds for the case of (DOP) orthogonality.

Consider the situation of rebalancing a portfolio of three securities; the feasible region is shown in the following set of plots. The projections were chosen to provide the best sense of the geometry involved. As with Figure 5.3, the expected return constraint is not shown for presentation purposes. First, consider the simplifying case of zero transaction costs. This case is illustrated by Figures 6.1 and 6.2. Each face is uniquely identified by color.

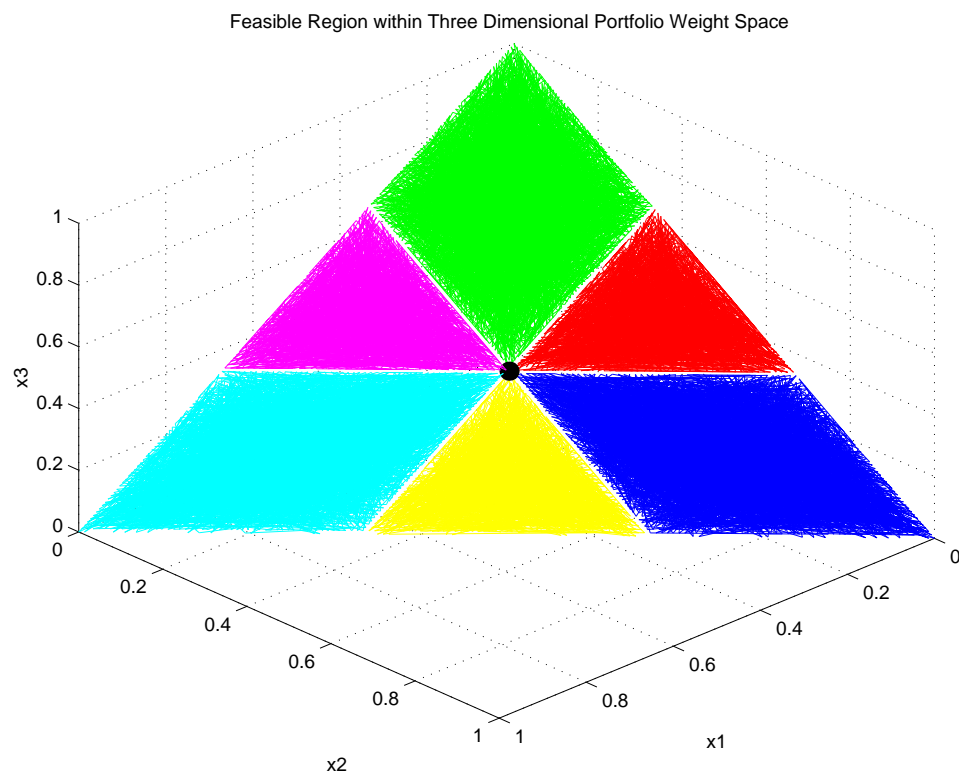


Figure 6.1: Without transaction costs, the problem is convex and the feasible region is just the portion of the $x_1 + x_2 + x_3 = 1$ plane in the nonnegative orthant. The example chosen is highly symmetric with an equal initial allocation located by the black dot.

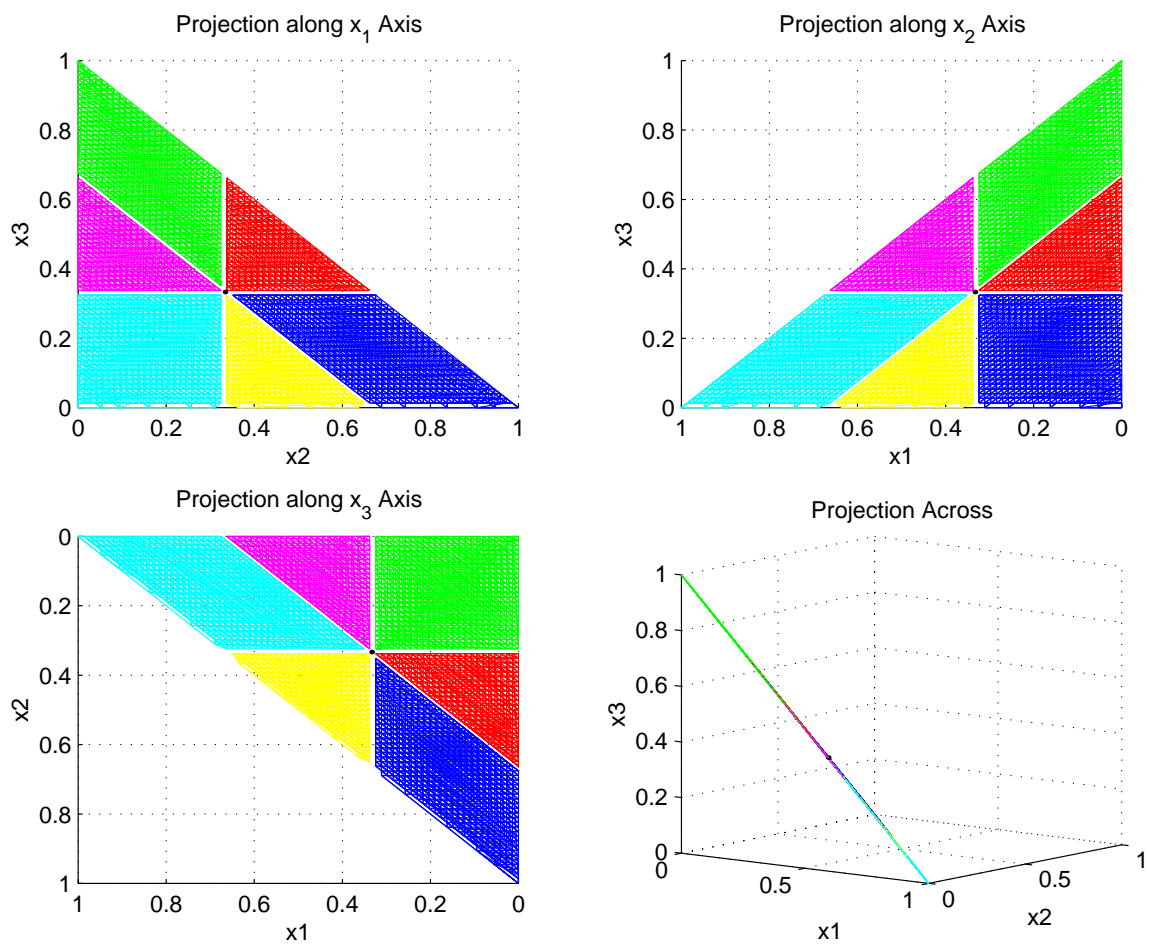


Figure 6.2: Projections along each axis and across are shown. The lower right panel clearly illustrates that this feasible region is planar.

Without transaction costs, the feasible region is convex and the entire machinery we have developed to solve (QPO) is not necessary. However, that case does set up a useful contrast. Figures 6.3 and 6.4 present similar perspectives of a feasible region for a three security portfolio. However, nonzero transaction costs are now involved.

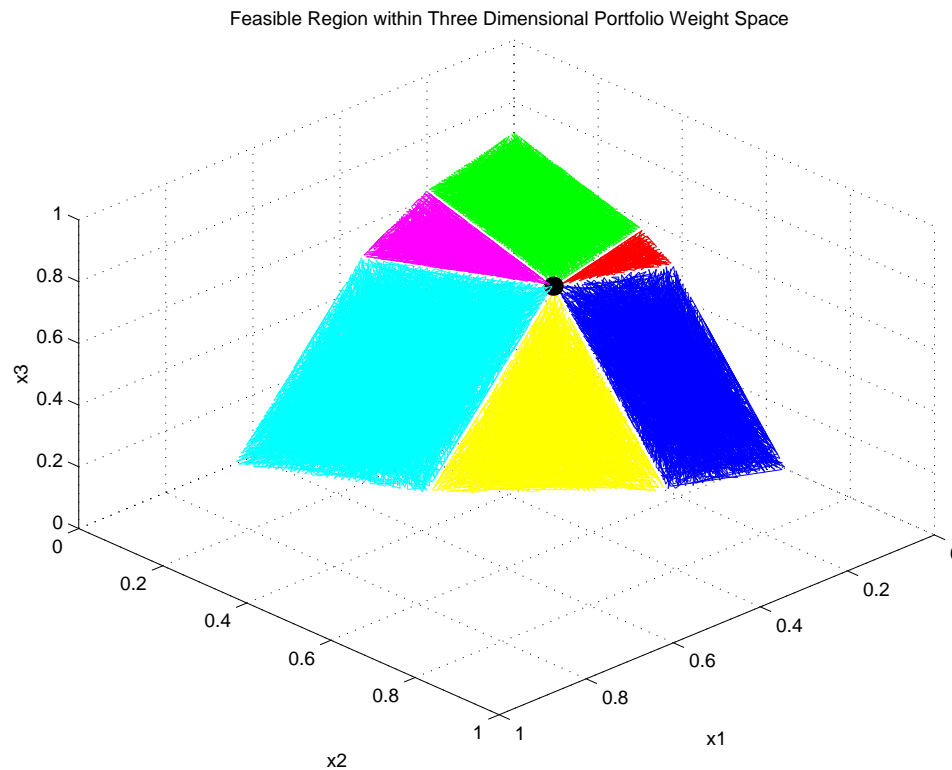


Figure 6.3: To provide orientation, the initial portfolio is again located by a black dot at the intersection. It is the only point remaining on the original plane and is the only point shared by all six faces. Here $\bar{x} = [.2, .3, .5]^T$, $c^B = [.2, .3, .4]^T$ and $c^S = [.5, .4, .3]^T$. The costs used are exaggerated for presentation purposes.

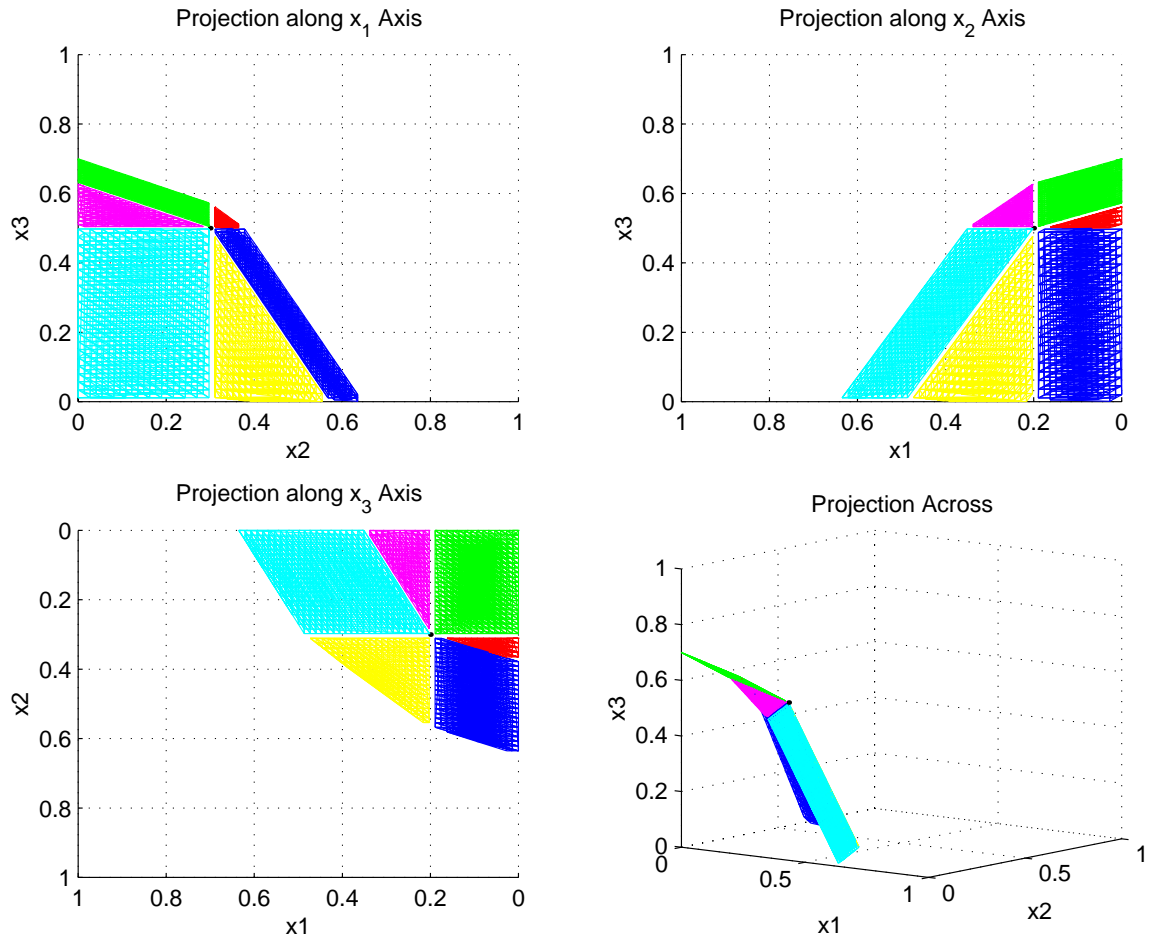


Figure 6.4: Again multiple projections are shown. Each face is uniquely identified by color to aid identification among the multiple perspectives. The lower right panel across the feasible region clearly shows the “umbrella-like” nonconvexity that is present.

Notice that with nonzero costs, the feasible region taken as a single entity is no longer convex. However, it “breaks” into separate faces each of which is a convex region. This is the geometric insight into why each QP-subproblem is straightforward to solve. Though there may be many of these QP-subproblems to consider, they are each convex problems.

The correspondence between active sets, QP-subproblems and faces can be explained in the following way. Imposing an active set means selecting its corresponding face in the feasible region and then solving its specific QP-subproblem. The optimal solution for that QP-subproblem locates a point within that face or perhaps on one of its decision boundaries. The later occurrence leads to a geometric interpretation of degeneracy and the discussion of Section 6.3.

6.3 Identifying the Neighboring Faces

Continuing with our three dimensional discussion, Figure 6.5 on the following page is an idealization of this geometry. On one level, this idealization can be understood simply because it looks like one of the projections seen earlier in this chapter. However, a deeper understanding of this idealization requires that several issues be addressed.

First, degeneracy in a solution to an instance of the portfolio rebalancing problem can be defined by the analytic condition: $\exists i$ s.t. $u_i = v_i = 0$. Also, the transformation introduced in Chapter 5 which relates the buying and selling variables to the space of portfolio weights was $x = \bar{x} + u - v$. Therefore in our idealized figure, the boundaries between faces are given by $x_i = \bar{x}_i$ for $i = 1, 2, 3$. A decision of activity corresponds geometrically to the selection of all faces to one side of such a decision boundary.

For Figure 6.5, suppose that decisions regarding x_1 and x_2 were originally made following solution of (rSDP). Intersecting these decisions identifies the two faces which must be considered. The best observed solution in this set of QP-subproblems is located by the point A.

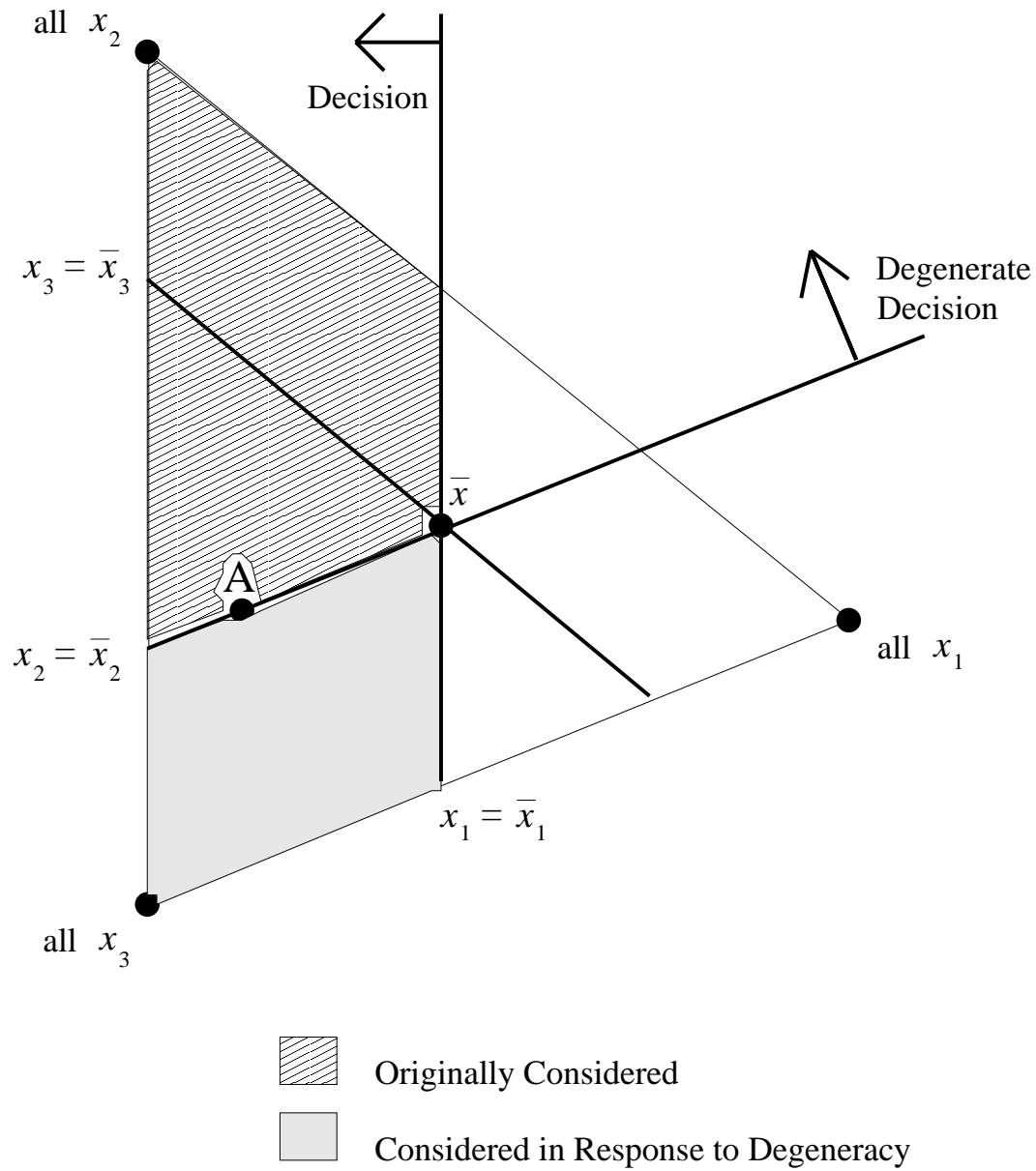


Figure 6.5: This figure idealizes the feasible region of the three security rebalancing problem. Point A is a degenerate solution which could be optimal. However, the additional neighboring face must be examined before that could be decided with any confidence.

By inspection, point A contains a degenerate decision for x_2 . Geometrically, this is recognizable since point A lies on the decision boundary $x_2 = \bar{x}_2$. The most conservative response is to simply “unmake” the degenerate decision for x_2 . This would mean that the only enforceable decision boundary concerns x_1 . Being to the left of the decision boundary $x_1 = \bar{x}_1$ means that one additional subproblem corresponding to the additional shaded face must be examined. Geometrically, that additional subproblem corresponds to a neighboring face since it shares a boundary.

Notice that the dimensionality of the problem makes a different type of representation necessary. For even two securities, the space of buying and selling decisions is already four-dimensional and not readily visualized. This was one reason we chose to establish the geometric idea of neighboring faces using the portfolio weight space of the three security rebalancing problem. Having demonstrated the geometric rationale for responding to degeneracy with the consideration of neighboring faces, we will transition to the alternate representation of trees.

6.4 Generality of the Tree Representation

The tree representation for enumerative algorithms was originally introduced in Section 4.5. The advantage of switching to the tree representation here is that it generalizes for higher dimensional problems. This generality will be important to the counting arguments that will be presented in later sections. Figure 6.6 presents a tree of total depth n_{op} where movement down into the tree is measured by the number of activity decisions made, n_a , and movement back up into the tree is measured by the number of degenerate decisions, n_d . Of course, $0 \leq n_d \leq n_a \leq n_{op}$ must be true.

In the previous section, the additional faces considered in response to degeneracy were neighboring in the natural, geometrical sense of the word. As will be illustrated, the idea can be carried over to the tree representation. In that context, “neighboring” QP-subproblems share a common path down through some depth of the tree to a branch where they diverge.

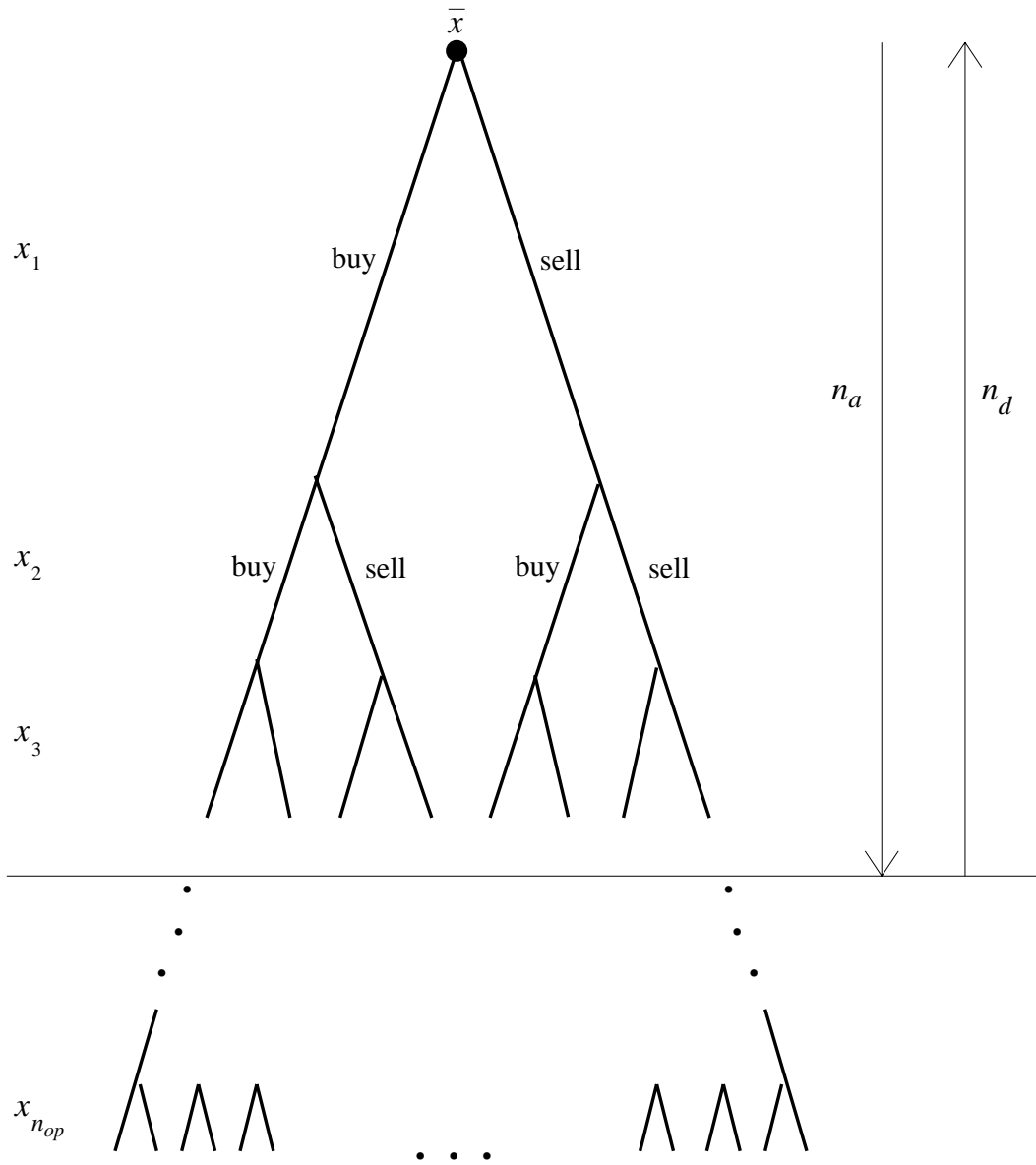


Figure 6.6: This presents the tree representation we will be using to describe our response to degeneracy. The natural ordering of decisions is used to simplify presentation though our computational implementation handles a general set of decisions.

To finish off our three dimensional example, Figure 6.7 on the following page presents the corresponding tree for the example of Figure 6.5. The same shading scheme is used and the QP-subproblem which gave rise to solution A is labeled. One observation is that although there are $2^3 = 8$ possible active sets, there are only 6 faces in the feasible region. The explanation is that a possible active set exists for each possible buy and sell combination. However, two of these active sets are immediately known to be infeasible. The infeasible active sets correspond to buying every security or selling every security. This cannot be consistent with the full investment constraint and has no geometrical realization. As a technical aside, the response to the degenerate decision regarding x_2 brings one of those infeasible QP-subproblems into consideration.

6.5 Ways to Grow the Search Space

There are multiple ways that additional QP-subproblems can be generated in response to degeneracy. The approach which has already been discussed in this chapter is conservative in the sense that it “unmakes” any decision that is found to be degenerate. This means that the number of additional subproblems grows exponentially each time a degenerate decision is found. Other approaches are more aggressive in the sense that fewer additional QP-subproblems are considered in response to degeneracy. The two approaches we will present here both rely on decision-making criteria which are similar to the one introduced back in Section 4.6.2. However for the recursive degeneracy check, the best observed QP-subproblem solution vector is supplied as the estimate x^* .

Before beginning our presentation of two specific degeneracy responses, an implicit characteristic of any QP-subproblem solution should be made explicit. Notice that the solution of every QP-subproblem has already had orthogonality imposed. That is to say there are only two possible estimates for each orthogonal pair: either $(x_i^*, 0)$ or $(0, x_j^*)$ where $x_i^*, x_j^* \geq 0$ depends on which variable was assumed to be active for that subproblem. This allows us to construct a specialized decision criteria for each approach.

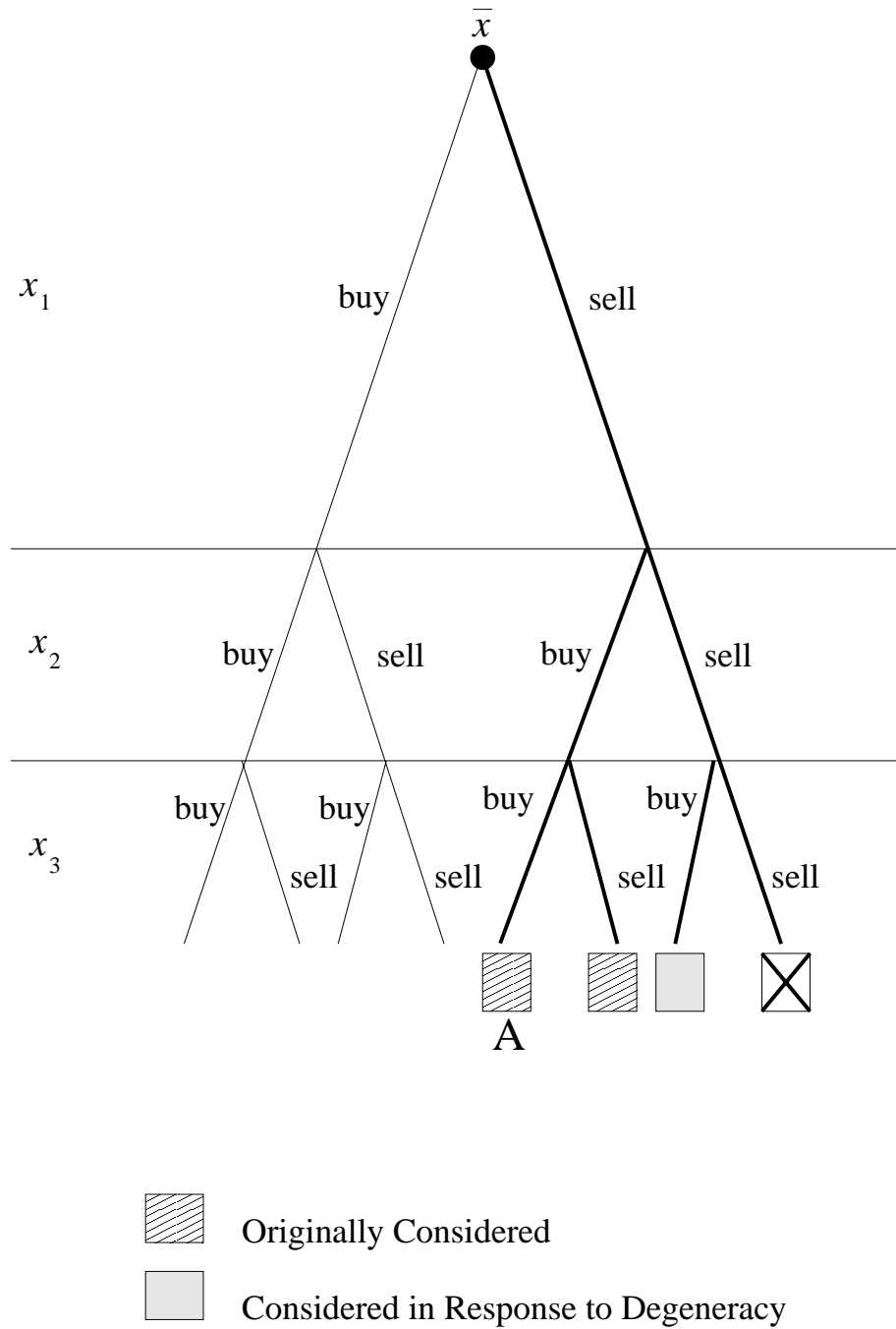


Figure 6.7: This figure shows the tree corresponding to Figure 6.5. Again, 2 subproblems are initially considered. In response to degeneracy, additional neighboring subproblems are considered. Bold lines indicate which branches were followed and connect to the leaves that were actually considered.

6.5.1 The Conservative Choice

The first approach is the most conservative way to grow the tree. It reverses any decision that is found to be degenerate and so doubles the number of QP-subproblems for each degenerate decision. Suppose an initial activity decision was made for the orthogonal pair (x_i, x_j) . The conservative decision criteria can be expressed in the following manner:

If x_i had been considered active but $x_i^* \approx 0$
 then x_j should now be considered active.

If x_j had been considered active but $x_j^* \approx 0$
 then x_i should now be considered active.

The analytic statement $x_i^* \approx 0$ was implemented as $x_i^* \leq \epsilon$ where ϵ is a numerical tolerance, for example $\epsilon = 10^{-8}$.

In terms of performance, the number of original QP-subproblems considered with this approach is $2^{(n_{op}-n_a)}$. If the best observed solution from this set is degenerate then the conservative degeneracy response is to consider additional, neighboring subproblems. Given n_d degenerate decisions, the total number of QP-subproblems that the conservative approach considers is: $2^{(n_{op}-n_a+n_d)}$.

As an aside, the naïve approach of Section 4.3.1 always generates $2^{n_{op}}$ subproblems. Remembering that $0 \leq n_d \leq n_a \leq n_{op}$, the conservative approach is considering that number times a factor less than unity: $2^{n_{op}} \times \{2^{(n_d-n_a)}\}$. That factor is a possible performance metric for this approach.

6.5.2 A More Aggressive Approach

Once an optimal solution is found for any QP-subproblem, several useful pieces of information are known. First, that QP-subproblem objective value is an upper bound on the optimal value of (QPO) just as (rSDP) was a lower bound on (QPO). Second -and of more direct interest- that solution vector constitutes an estimate of the complete optimal active set. Even with orthogonal pairs of variables for which no activity decision was made, decisions regarding what was bought and what was sold

can be observed. The previous conservative approach discarded that information. This more aggressive approach makes use of that information.

Again, let x^* be the best observed QP-subproblem optimal solution vector. However, suppose no initial activity decision was made for the orthogonal pair (x_i, x_j) . The aggressive approach also uses (x_i^*, x_j^*) as its estimate and applies the following decision criteria:

If $\frac{x_i^*}{u_i} \geq U_{tol}$ then x_i should now be considered active.

If $\frac{x_j^*}{u_j} \geq U_{tol}$ then x_j should now be considered active.

Here, U_{tol} is the same user-specified tolerance from Section 4.6.2. It is important to stress that nondegenerate decisions are never reversed by this criteria.

In simplest terms, the aggressive approach keeps all decisions which can be inferred from the best observed solution. Figure 6.8 and 6.9 on subsequent pages show in sequence an example of how this approach works. The analytic expression operative for the aggressive approach is $2^{(n_{op}-n_a)} + \{2^{n_d} - 1\}$. The growth here is additive which is slower than exponential growth. So, fewer additional QP-subproblems will be enumerated in response to degeneracy.

6.6 Effect of NP-hardness on the Strategy

To conclude this discussion of our response to degeneracy, it is appropriate to again emphasize that we are considering an NP-hard optimization problem. So we cannot guarantee that our strategy with either a conservative or aggressive degeneracy response will always identify the optimal solution. Being aggressive by considering fewer subproblems is a systematic way to reduce the amount of computation but does not guarantee better performance. Being conservative systematically increases the number of subproblems brought into consideration but still does not provide a guarantee of optimality. However, in real-world situations where suboptimal solutions offering incremental improvement have their own value, this variety in approach is a useful feature of our strategy.

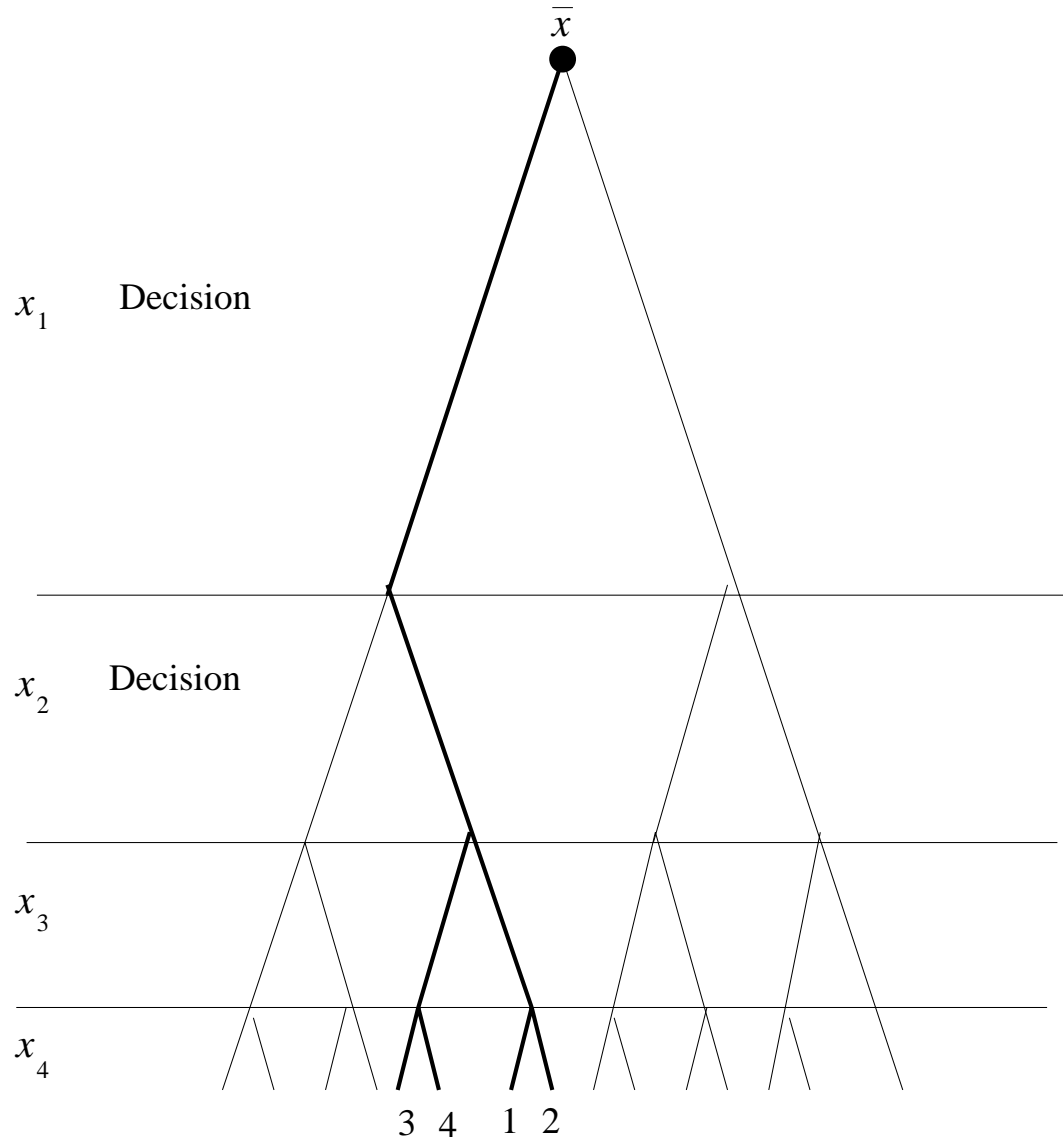


Figure 6.8: This figure presents the example of a four layer tree in which two decisions are initially made following the solution of (rSDP). Again, bold lines indicate where decisions were made and which leaves are actually considered. The observed solutions are ranked by objective value. The best observed solution is numbered 1. Assume that x_3 and x_4 take on optimal values within $U_{tol}\%$ of their respective upper bounds.

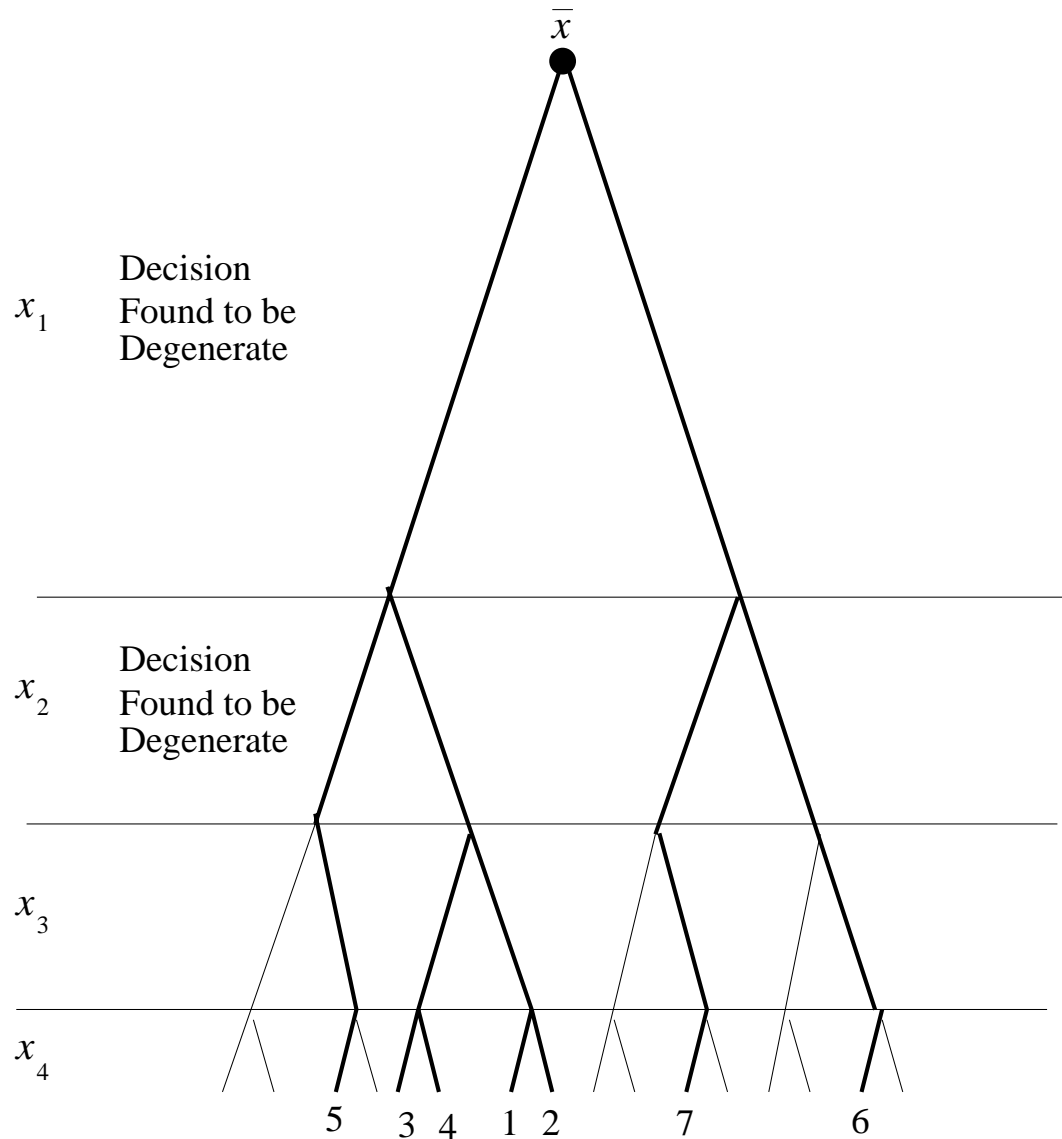


Figure 6.9: This figure shows the aggressive response to degeneracy. Since both decisions were found to be degenerate then the conservative approach would force consideration of all 16 leaves. Using the inferred nondegenerate decisions from solution #1, a branching structure for x_3 and x_4 can be translated to select only subproblems #5, 6, and 7. Notice that with the aggressive approach only 7 of the 16 subproblems are considered.

CHAPTER 7

Final Discussion and Concluding Remarks

Here we take the opportunity to summarize our work and highlight the main contributions of this thesis. We conclude by briefly introducing several areas where continued research could occur.

7.1 Summary

This thesis was inspired by a topic drawn from financial optimization. The portfolio rebalancing problem with our percentage cost model necessitated the introduction of nonconvex orthogonality constraints. From that specific formulation, a new general type of mathematical programming problem was considered. The name given to this general type of problem was (QPO), a quadratic programming problem subject to orthogonality constraints. The fact that (QPO) fell under the Q^2P classification originally motivated our research into the applicability of semidefinite programming to (QPO).

In Chapters 2 & 3, we documented several of the theoretical hurdles that were crossed in order to successfully construct a general semidefinite relaxation of (QPO). Imposing orthogonality, homogenizing and then lifting linear constraints, identifying and imposing the correct bounds: all were addressed.

Solving that semidefinite relaxation (rSDP) as the first stage in an enumerative algorithm was found to be a successful strategy for the portfolio rebalancing problem. From computational experiments, we realized that to make the enumerative algorithm more robust, an understanding and response for degeneracy was required. Chapter 6 showed geometrically that an appropriate response to degeneracy was the consideration of neighboring QP-subproblems.

7.2 Main Contributions

In the course of this research, a wide range of mathematical programming topics were applied and in some cases extended. Financial insights helped to moti-

vate the cost model as well as identify an appropriate risk measurement. It is those extensions and insights that are the main contributions of this thesis.

7.2.1 Finding a New Role for y

Mathematically, the use of a scalar variable to incorporate the linear term into a quadratic objective can be found elsewhere. However, the continued use of that same variable to homogenize any linear equalities appears to be a new and novel contribution of this thesis. The clearest presentation of this idea is found in the following formulation of (QPO):

$$\begin{aligned}
 \min_{x,y} \quad & \begin{bmatrix} x^T & y \end{bmatrix} \begin{bmatrix} Q & \frac{c}{2} \\ \frac{c^T}{2} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
 \text{s.t.} \quad & Ax - by = 0 \\
 & y = 1 \\
 & 0 \leq x \leq u \\
 & x_i x_j = 0 \quad \forall (i, j) \in \Phi, \text{ given acceptable } l.
 \end{aligned}$$

7.2.2 The Symmetric Lifting Procedure

In the course of constructing our semidefinite relaxation, the most complicated issue addressed was how to lift the linear equalities on the vector variable z and impose them on the matrix variable Z . The most compact representation of the lifting procedure we developed is:

$$A_k = a_l^T e_j^T + e_j a_l \text{ for } j = 1, \dots, n \text{ and } l = 1, \dots, m.$$

This expression clearly relates the $m \times n$ linear system $Az = 0$ to the set of mn matrix constraints each having the form $A_k \bullet Z = 0$.

The two specific contributions shown above allowed us to successfully use semidefinite programming as an alternative to the traditional MIP approach for combinatorial problems involving continuous variables. The use of semidefinite programming in this way is probably the biggest conceptual contribution of this thesis.

7.2.3 Scaling the Risk Measurement

The final contribution we would like to highlight is the inclusion of transaction costs as an $(n + 1)^{st}$ security and the scaled risk measurement that was constructed. Refreshing the reader, that fractional objective was:

$$\frac{\frac{1}{2}x^T V x}{\{1 - x_{n+1}\}^2}$$

This objective scales the standard risk measurement by the square of the dollar amount actually invested. Without transaction costs, this theoretical extension recovers the standard portfolio risk measurement. Appendix C describes the mathematical transformation that was used to accommodate this fractional objective.

7.3 Directions for Future Work

No research program is ever complete. At the outset, research is performed to answer a certain set of questions. However, answers to those original questions always lead to additional questions. To conclude this thesis, we point out some of the most interesting areas where further research could be performed.

7.3.1 Multi-period Rebalancing

Returning to an aside made in Section 5.6, the entire discussion of Chapter 5 exists within the context of a single rebalancing event for a portfolio. However, investment portfolios are often managed from a long-term perspective. The time horizon of an individual investor may be 20, 30, or even 40 years. For institutions, the appropriate time horizon may be still longer. Over such lengthy periods of time, there would certainly be many opportunities to at least consider the possible value of rebalancing a portfolio. One clear research direction would be to incorporate the solution strategy for a single rebalancing decision into a multi-period scheme. In such a context, issues such as the optimal rebalancing frequency could be studied.

7.3.2 Explicitly Constraining the Costs Paid to Rebalance

A general theme in optimization is the inter-relationship between the constraints and the objective. In Chapter 5, we modified our objective function to take into account the increased riskiness of transaction cost depleted portfolios. In effect, we are penalizing small denominators in our fractional objective.

However, one can imagine imposing an explicit constraint on the denominator itself. Financially, such a constraint would place a limit on the amount the investor is willing to pay in transaction costs in order to rebalance their portfolio. This constraint could have the form $1 - x_{n+1} \geq \alpha$ where α is an additional parameter. This seems to be a realistic constraint which could be imposed without raising complications in our solution strategy.

7.3.3 Pursuing Additional Applications

Ideas in this area fall into two broad categories: (1) identify other applications for (QPO) and (2) identify other uses for our solution strategy.

In the portfolio rebalancing application, the orthogonality constraint we needed to impose was simply “buy OR sell, don’t do both”. That is a very straightforward qualification, suitable for many other applications. As an exercise, we have already stated several instances of a math programming problem with equilibrium constraints, (MPEC), in (QPO) standard form. One such instance was a bi-level optimization problem [6]. As we mentioned in Chapter 1, (QPO) may be a productive alternative for approaching these types of problems.

In addition to finding applications for (QPO) in other disciplines, our solution strategy has applications elsewhere within mathematical programming. Specifically, our (rSDP) approach could be run as a preprocessing step for the traditional (MIP) approach. We use the results of (rSDP) to make activity decisions -where possible- before enumerating all remaining possibilities. However, the results of (rSDP) could instead be used to fix some subset of binary variables in a MIP formulation. Where activity decisions are not possible, the corresponding binary variables are simply left free. The reader is referred to Section 1.3.1 for an example of a MIP.

7.3.4 Improving the Implementation

Finally, the last area where future work could occur concerns the continued development and refinement of the computational implementation. For example, the solution strategy we presented for the case of (DOP) orthogonality is readily generalizable for (SOS1) orthogonality. Other extensions may be possible.

Back in Section 4.6.2, we promised to describe a heuristic which is perhaps better suited for larger problems. This alternative is to rank the decisions and always generate QP-subproblems corresponding to the weakest N decisions where N is specified by the user. This single parameter N takes the place of the two decision tolerances, L_{tol} and U_{tol} . This idea has the computational advantage that the number of QP-subproblems is fixed. Such control over subproblem generation could perhaps be useful for balancing the computational load if subproblems are solved in parallel.

This leads us to the final point we will discuss. Parallel computation offers obvious benefits for our solution strategy. Speaking in terms of the portfolio re-balancing problem, all QP-subproblems and even each instance of (rSDP) along a TCEF could be solved in parallel. That could significantly reduce the time needed to compute a complete TCEF. Parallel computation would help alleviate the computational “bottleneck” that is created by large numbers of QP-subproblems.

LITERATURE CITED

- [1] S. P. Aggarwal. Indefinite quadratic fractional programming with a quadratic constraint. *Cahiers du Centre études de recherche opérationnelle*, 15:405–410, 1973.
- [2] S. P. Aggarwal. Quadratic fractional functionals programming. *Cahiers du Centre études de recherche opérationnelle*, 15:157–165, 1973.
- [3] F. Alizadeh, J.-P. Haeberly, M. V. Nayakkankuppam, M. L. Overton, and S. Schmieta. SDPpack user’s guide. Technical report, New York University Computer Science Department, New York, NY, June 1997.
- [4] F. Andersson, H. Mausser, D. Rosen, and S. Uryasev. Credit risk optimization with conditional value-at-risk criterion. *Mathematical Programming*, 89(2):273–309, 2001.
- [5] K. M. Anstreicher. A monotonic projection algorithm for fractional linear programming. *Algorithmica*, 1:483–498, 1986.
- [6] Jonathan F. Bard. Convex two-level optimization. *Mathematical Programming*, 40:15–27, 1986.
- [7] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley, New York, second edition, 1993.
- [8] D. Bertsimas, C. Darnell, and R. Soucy. Portfolio construction through mixed-integer programming at Grantham, Mayo, Van Otterloo and Company. *Interfaces*, 29(1):49–66, 1999.
- [9] B. Borchers. CSDP 2.3 User’s Guide. *Optimization Methods and Software*, 11:597–611, 1999.
- [10] S. Burer and R. D. C. Monteiro. An efficient algorithm for solving the maxcut sdp relaxation. Technical report, School of ISyE, Georgia Tech, Atlanta, GA, December 1998.
- [11] T.-J. Chang, N. Meade, J. E. Beasley, and Y. M. Sharaiha. Heuristics for cardinality constrained portfolio optimisation. Technical report, The Management School, Imperial College, London SW7 2AZ, England, May 1998.
- [12] A. Charnes and W. W. Cooper. Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 9:181–186, 1962.

- [13] A. Consiglio and S. Zenios. Integrated simulation and optimization models for tracking international fixed income indices. *Mathematical Programming*, 89(2):311–339, 2001.
- [14] I. R. de Farias Jr., E.L. Johnson, and G.L. Nemhauser. Branch-and-cut for combinatorial optimization problems without auxiliary binary variables. Technical report, Department of Industrial Engineering, SUNY Buffalo, February 2000.
- [15] J. W. Demmel. *Applied Numerical Linear Algebra*. SIAM, Philadelphia, PA, 1997.
- [16] I. R. de Farias Jr., E. L. Johnson, and G. L. Nemhauser. A generalized assignment problem with special ordered sets: a polyhedral approach. *Mathematical Programming*, 89(1):187–203, 2000.
- [17] R. W. Freund and F. Jarre. An interior-point method for fractional programs with convex constraints. *Mathematical Programming*, 67:407–440, 1994.
- [18] R. W. Freund and F. Jarre. An interior-point method for multifractional programs with convex constraints. *Journal Of Optimization Theory And Applications*, 85:125–161, 1995.
- [19] R. W. Freund and F. Jarre. Solving the sum-of-ratios problem by an interior-point method. Technical Report Numerical Analysis Manuscript 99 3 13, Bell Laboratories, Murray Hill, New Jersey, June 1999.
- [20] Michel Goemans and Franz Rendl. Combinatorial optimization. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, chapter 12, pages 343–360. Kluwer Academic Publishers, Boston, MA, 2000.
- [21] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42:1115–1145, 1995.
- [22] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 1983.
- [23] C. Helmberg. *SBmethod: A C++ Implementation of the Spectral Bundle Method, version 1.1*, 2000.
- [24] C. Helmberg and K. Kiwiel. A spectral bundle method with bounds. Technical Report SC-99-37, Konrad-Zuse-Zentrum fuer Informationstechnik, Berlin, December 1999.
- [25] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM Journal on Optimization*, 10(3):673–696, 2000.

- [26] H. Konno and A. Wijayanayake. Portfolio optimization problem under concave transaction costs and minimal transaction unit constraints. *Mathematical Programming*, 89(2):233–250, 2001.
- [27] K. Krishnan and J. E. Mitchell. A linear programming approach to semidefinite programming. Technical report, Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, May 2001.
- [28] Serge Kruk and Henry Wolkowicz. General nonlinear programming. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, chapter 20, pages 563–575. Kluwer Academic Publishers, Boston, MA, 2000.
- [29] E. K. Lee and J. E. Mitchell. Computational experience of an interior point SQP algorithm in a parallel branch-and-bound framework. In H. Frenk *et al.*, editor, *High Performance Optimization*, chapter 13, pages 329–347. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.
- [30] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, 1996.
- [31] Harry M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Blackwell Publishers, 2nd edition, 1991.
- [32] The MathWorks Inc. *Using MATLAB*. The MathWorks Inc., Natick, MA.
- [33] H. D. Mittelmann. An independent benchmarking of SDP and SOCP solvers. Technical report, Arizona State University, Department of Mathematics, June 2001.
- [34] J. M. Mulvey. Introduction to financial optimization: Mathematical Programming special issue. *Mathematical Programming*, 89(2), 2001.
- [35] G. L. Nemhauser and L. A. Wolsey. *Integer and Combinatorial Optimization*. John Wiley, New York, 1988.
- [36] A. Nemirovski, C. Roos, and T. Terlaky. On maximization of quadratic form over intersection of ellipsoids with common center. *Mathematical Programming*, 86(3):463–473, 1999.
- [37] Yu. E. Nesterov. Global quadratic optimization via conic relaxation. Technical report, Catholic University of Louvain, Belgium, February 1998.
- [38] Yuri Nesterov, Henry Wolkowicz, and Yinyu Ye. Nonconvex quadratic optimization. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, chapter 13, pages 361–419. Kluwer Academic Publishers, Boston, MA, 2000.

- [39] Pablo A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. Technical report, Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, 2001.
- [40] F. Potra, R. Sheng, and N. Brixius. SDPHA – a MATLAB implementation of homogeneous interior-point algorithms for semidefinite programming. Technical report, Department of Mathematics, University of Iowa, Iowa City, Iowa 52242, April 1997. Version 3: March 1998.
- [41] S. Schaible. Fractional programming. In R. Horst and P. Pardalos, editors, *Handbook of Global Optimization*, pages 495–608. Kluwer Academic Publishers, Dordrecht, The Netherlands, November 1994.
- [42] M. C. Steinbach. Markowitz revisited: mean-variance models in financial portfolio analysis. *SIAM Review*, 43(1):31–85, 2001.
- [43] M. J. Todd. Semidefinite optimization. Technical report, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, February 2001.
- [44] K. C. Toh, M. J. Todd, and R. Tütüncü. *SDPT3 — a Matlab software package for semidefinite programming, version 2.1*, 1999.
- [45] K. C. Toh, M. J. Todd, and R. Tütüncü. *SDPT3 — a Matlab software package for semidefinite programming, version 3.0*, 2001.
- [46] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1996.
- [47] H. Wolkowicz. Simple efficient solutions for semidefinite programming. Technical Report CORR 2001–49, Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, N2L 3G1 Canada, September 2001.
- [48] Y. Ye. Approximating quadratic programming with bound and quadratic constraints. *Mathematical Programming*, 84:219–226, 1999.
- [49] Y. Ye and S. Zhang. New results on quadratic minimization. Technical report, Department of Management Sciences, The University of Iowa, Iowa City, Iowa 52242, April 2001.
- [50] S. Zhang. Quadratic minimization and semidefinite relaxation. *Mathematical Programming*, 87(3):453–465, 2000.
- [51] Y. Zhao and W. T. Ziemba. A stochastic programming model using an endogenously determined worst case risk measure for dynamic asset allocation. *Mathematical Programming*, 89(2):293–309, 2001.

APPENDIX A

Portfolio Datasets Used

As previously mentioned, the computational results of Chapter 5 used two sources of historical data. Additional information regarding those datasets is contained in this appendix.

A.1 Markowitz Dataset

Company Name	Security Reference Number Used	Average Annual Return (%)	Standard Deviation (%)
American Tobacco	1	6.59	23.11
AT&T	2	6.16	12.12
United States Steel	3	14.61	29.24
General Motors	4	17.34	30.90
Atchison, Topeka & Sante Fe	5	19.81	35.76
Coca-Cola	6	5.51	20.31
Borden	7	12.76	16.98
Firestone	8	19.03	38.31
Sharon Steel	9	11.56	28.15

Table A.1: The reference number is the component number of the portfolio weight vector. For example, the 5th security has the highest individual return. The ensemble average performance was an annual return of 12.60% with a standard deviation of 26.10%.

A.2 Dow Jones Dataset

This dataset was constructed using the past 8 years of price history for the 30 stocks which make up the Dow Jones Industrial Average as of the writing of this thesis. This time frame spanned one of the greatest bull markets in American history. Overlapping annual returns were calculated during this period to increase the number of data points used to calculate the covariance matrix.

Company Name (Ticker Symbol)	Security Reference Number Used	Average Annual Return (%)	Standard Deviation (%)
ALCOA Inc. (AA)	1	24.04	30.84
AlliedSignal Inc. (ALD)	2	12.28	23.69
American Express Co. (AXP)	3	30.13	26.52
Boeing Co. (BA)	4	19.37	30.93
Citigroup Inc. (C)	5	44.86	31.28
Caterpillar Inc. (CAT)	6	10.79	26.94
Dupont Co. (DD)	7	9.70	25.91
Walt Disney Co. (DIS)	8	13.75	23.61
Eastman Kodak Co. (EK)	9	0.02	23.48
General Electric Co. (GE)	10	34.51	19.18
General Motors Corp. (GM)	11	6.63	18.50
Home Depot (HD)*	12	34.51	35.47
Hewlett Packard Co. (HWP)	13	26.32	39.78
IBM Corp. (IBM)	14	35.31	33.81
Intel (INTC)*	15	49.43	56.92
International Paper Co. (IP)	16	0.94	19.87
Johnson & Johnson (JNJ)	17	23.15	18.59
JP Morgan & Co. (JPM)	18	25.51	28.39
Coca-Cola Co. (KO)	19	13.99	26.36
McDonalds Corp. (MCD)	20	13.56	27.26
Minn. Mining & Man. Co. (MMM)	21	13.05	18.56
Phillip Morris Co. (MO)	22	19.17	42.36
Merck & Co. (MRK)	23	26.17	21.99
Microsoft (MSFT)*	24	47.18	46.48
Procter & Gamble Co. (PG)	25	16.42	26.80
SBC Communications, Inc. (SBC)*	26	15.59	22.02
AT&T Corp. (T)	27	2.02	40.12
United Technologies Corp. (UTX)	28	29.08	21.49
Walmart (WMT)	29	31.69	35.72
ExxonMobil Corp. (XOM)	30	18.19	12.71

Table A.2: The reference number is again the vector component number. For example, the 15th security has the highest individual return. Those marked with an asterisk are the most recent additions to the index. The ensemble average performance was an annual return of 21.58% with a standard deviation of 28.52%. As an aside, notice that the ensemble average of the securities added to the index was much higher at 36.68%.

APPENDIX B

Bounding the Portfolio Expected Return

As mentioned in Chapter 5, an efficient frontier is generated by solving multiple instances of (QPO) where the expected portfolio return parameter is varied over an interval. However, identifying the endpoints of that interval (ie. the minimum and maximum possible returns) for a transaction cost efficient frontier is a nontrivial exercise.

B.1 Maximum Portfolio Return

Notice that since $x = \bar{x} + u - v$ then the portfolio return is expressible as:

$$\mu^T x = \mu^T (\bar{x} + u - v) = \mu^T \bar{x} + \mu^T (u - v).$$

Dropping the constant term, this is the linear objective that should be maximized subject to full investment, bound, and orthogonality constraints. The problem of interest is:

$$\begin{aligned} \max_{u,v} \quad & \begin{bmatrix} \mu^T & -\mu^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ \text{s.t.} \quad & \begin{bmatrix} (e_n + c^B)^T & (c^S - e_n)^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = 0 \\ & 0 \leq u_i \leq 1 - \bar{x}_i \\ & 0 \leq v_i \leq \bar{x}_i \\ & u_i v_i = 0 \quad \forall i \end{aligned}$$

Though possible, we do not need to apply our (rSDP) solution strategy for this problem. Instead, we use a greedy algorithm that is very similar to the test problem presented in Section 4.7.2. In words, orthogonality can be imposed simply by selecting appropriate entries from a sorted list. However, it is not possible to identify the minimum efficient portfolio return with an analogous linear programming problem.

B.2 Minimum Portfolio Return

Therefore to locate the minimum efficient expected return we use the (rSDP) solution strategy to solve the following instance of (QPO):

$$\begin{aligned}
 \min_{u,v} & \left[\bar{x}^T V \quad : \quad -\bar{x}^T V \right] \begin{bmatrix} u \\ v \end{bmatrix} + \frac{1}{2} \begin{bmatrix} u^T & v^T \end{bmatrix} \begin{bmatrix} V & -V \\ -V & V \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\
 \text{s.t.} & \begin{bmatrix} (e_n + c^B)^T & (c^S - e_n)^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = 0 \\
 & 0 \leq u_i \leq 1 - \bar{x}_i \\
 & 0 \leq v_i \leq \bar{x}_i \\
 & u_i v_i = 0 \quad \forall i
 \end{aligned}$$

Here the variance is minimized subject only to the full investment, bound, and orthogonality constraints. Scaling of the risk measurement as discussed in Section 5.4.3 can be accommodated as well. From the optimal solution of this problem, we can calculate the corresponding minimum expected return in a straightforward manner.

APPENDIX C

Transforming the Fractional Objective

Also in Chapter 5, we argued for financial and dimensional reasons that a fractional quadratic objective is a more appropriate choice for the portfolio rebalancing application. That fractional objective function consists of a quadratic numerator and the square of a linear function as the denominator. Given this mathematical structure, it is possible to transform our fractional objective function into an equivalent strictly quadratic objective.

Consider the following general problem:

$$\begin{aligned} \min \quad & \frac{\frac{1}{2}x^T Q x}{(h + g^T x)^2} \\ \text{s.t.} \quad & Ax = b \quad (FQP) \\ & x \geq 0, \end{aligned}$$

where Q is a symmetric, positive semidefinite matrix. Further, assume that the denominator satisfies $h + g^T x > 0$ if x is feasible in (FQP) — this holds if, for example, $h > 0$ and $g \geq 0$.

We change variables by letting $t \equiv \frac{1}{h + g^T x}$ and then defining $\hat{x} \equiv tx$. Note that we now have $g^T \hat{x} = 1 - ht$. The equality constraints $Ax = b$ are equivalent to $tAx = tb$, that is, $A\hat{x} = bt$.

Thus, our original problem (FQP) is equivalent to the following problem:

$$\begin{aligned} \min \quad & \frac{1}{2}\hat{x}^T Q \hat{x} \\ \text{s.t.} \quad & A\hat{x} - bt = 0 \quad (QP^*) \\ & g^T \hat{x} + ht = 1 \\ & \hat{x}, t \geq 0. \end{aligned}$$

Note that we really want solutions with $t > 0$. An appropriate lower bound for t can be found from any upper bound on $h + g^T x$. Once we find a solution (\hat{x}^*, t^*) , we can obtain a solution to the original problem by rescaling \hat{x} , so $x^* = \frac{1}{t^*}\hat{x}^*$.

This technique of replacing the denominator by a variable was used by Charnes and Cooper [12] for fractional programs where the objective is a ratio of linear functions and the constraints are linear. Bazaraa et al. [7] also discuss this approach. Other references on fractional programming include [1, 2, 5, 17, 18, 19, 41]. It is simply stated here that the presence of a linear term in the numerator can be accommodated with only slight modifications to this framework.