A Quasi-Newton Algorithm For Continuous Minimax

With Applications To Risk Management

In Finance

by

Melendres A. Howe

August, 1994

A thesis submitted for the degree of
Doctor of Philosophy of the University of London

Department of Computing
Imperial College of Science, Technology and Medicine
University of London
London SW7 2BZ
ABSTRACT

We present a superlinearly convergent quasi-Newton algorithm designed for continuous minimax problems. The algorithm uses a quadratic approximation to the objective function with a second order term that is conditioned by a Hessian approximation. We establish the global and local convergence properties of the algorithm; we show that it converges superlinearly, and that the maximizer of the quadratic approximation, augmented with a penalty term, corresponds to the minimum-norm subgradient. The use of the penalty term gives the algorithm a simple structure.

We apply the algorithm to several test problems to evaluate its performance. The solutions it finds are consistent with solutions found by other algorithms. In general, the performance of the proposed algorithm is consistent with its theoretical superlinear convergence property.

We formulate a minimax strategy for risk management in finance, and address the problem of hedging the risk of writing call options. We use the algorithm as a tool for implementing this strategy. We develop seven variants of the minimax strategy, three of which constrain transaction costs. In a large simulation and a limited empirical study, we measure the performance of the variants against a benchmark strategy, delta hedging, and test several hypotheses. The variants that constrain transaction costs consistently outperform delta hedging for a wide range of moneyness. These variants perform best when hedging at-the-money options, especially when the price of the underlying stocks oscillates about the exercise price.

We extend the minimax strategy into a multi-period setting, in which we consider different time frames, and to a CAPM-market-based setting, in which we consider portfolios of options. In simulation studies these extensions outperform the original minimax strategy.

We show that continuous minimax algorithms reach mid-range solutions when applied to hedging, and demonstrate that basic continuous minimax and extensions to it could be a basis for practical hedging strategies in the risk management of derivative securities. These results suggest that continuous minimax could make a unique contribution to programs of efficient risk management.
ACKNOWLEDGEMENTS

For whatever merit this dissertation may have I thank:

For a research studentship:
SERC

For providing a demanding academic environment:
Imperial College

For being an even more academically demanding supervisor:
Berc Rustem

For very extensive advice on many aspects of mathematical finance and hedging:
Michael Selby

For valuable insights into worst case scenarios in hedging:
Henrik Neuhaus

For advice on English, based on his linguametric research, some little of which I took, he being my husband:
James Howe

For being there:
my children, Hume and Sei
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENTS</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>TABLE OF CONTENTS</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>10</td>
</tr>
<tr>
<td>0</td>
<td>INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>A REVIEW OF CONTINUOUS MINIMAX</td>
<td>14</td>
</tr>
<tr>
<td>1.1</td>
<td>Introduction</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>Basic concepts in continuous minimax</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>The optimality condition</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>The differentiability of the max-function</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>The Hessian of the max-function</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>Failure of smooth methods</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>Methods to solve continuous minimax</td>
<td>19</td>
</tr>
<tr>
<td>8</td>
<td>Comments on Chaney's algorithm</td>
<td>19</td>
</tr>
<tr>
<td>9</td>
<td>Comments on Panin's algorithm</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>Comments on Kiwiel's algorithm</td>
<td>20</td>
</tr>
<tr>
<td>11</td>
<td>The relevance of global optimization algorithms</td>
<td>21</td>
</tr>
<tr>
<td>Appendix 1</td>
<td>Chaney's Algorithm</td>
<td>22</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>Panin's Algorithm</td>
<td>27</td>
</tr>
<tr>
<td>Appendix 3</td>
<td>Kiwiel's Algorithm</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>A QUASI-NEWTON ALGORITHM FOR CONTINUOUS MINIMAX</td>
<td>32</td>
</tr>
<tr>
<td>1.1</td>
<td>The continuous minimax problem</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>Definitions</td>
<td>35</td>
</tr>
<tr>
<td>2.1</td>
<td>Assumptions</td>
<td>36</td>
</tr>
<tr>
<td>2.2</td>
<td>Necessary Condition for an Extremum (nce)</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>The Algorithm</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>Justification for the Search Direction Subproblem</td>
<td>38</td>
</tr>
<tr>
<td>5.</td>
<td>Basic Convergence Results</td>
<td>42</td>
</tr>
<tr>
<td>6</td>
<td>Convergence Theorem and Rate of Convergence</td>
<td>53</td>
</tr>
<tr>
<td>Appendix</td>
<td>The existence of the Hessian of $\Phi(x)$</td>
<td>62</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3</td>
<td>NUMERICAL RESULTS</td>
<td>65</td>
</tr>
<tr>
<td>1</td>
<td>Introduction</td>
<td>65</td>
</tr>
<tr>
<td>1.1</td>
<td>Definitions</td>
<td>66</td>
</tr>
<tr>
<td>1.2</td>
<td>The Proposed Quasi-Newton Algorithm</td>
<td>68</td>
</tr>
<tr>
<td>1.3</td>
<td>Kiwiel's Algorithm</td>
<td>69</td>
</tr>
<tr>
<td>2</td>
<td>Numerical Results</td>
<td>72</td>
</tr>
<tr>
<td>2.1</td>
<td>Implementation, Output Variables and the Stopping Criterion</td>
<td>72</td>
</tr>
<tr>
<td>2.2</td>
<td>Convex-Convex Problems</td>
<td>74</td>
</tr>
<tr>
<td>2.3</td>
<td>Convex-Convex Problems</td>
<td>87</td>
</tr>
<tr>
<td>2.4</td>
<td>Discrete Minimax Problems</td>
<td>94</td>
</tr>
<tr>
<td>3</td>
<td>Conclusion</td>
<td>101</td>
</tr>
<tr>
<td>4</td>
<td>MINIMAX HEDGING STRATEGY</td>
<td>102</td>
</tr>
<tr>
<td>1</td>
<td>Introduction</td>
<td>102</td>
</tr>
<tr>
<td>1.1</td>
<td>Minimax: the hedging strategy and the algorithm</td>
<td>103</td>
</tr>
<tr>
<td>1.2</td>
<td>Notation</td>
<td>104</td>
</tr>
<tr>
<td>2</td>
<td>Introduction to options and the hedging problem</td>
<td>104</td>
</tr>
<tr>
<td>2.1</td>
<td>Stock options</td>
<td>104</td>
</tr>
<tr>
<td>2.2</td>
<td>Determinants of call premium</td>
<td>106</td>
</tr>
<tr>
<td>2.3</td>
<td>The hedging problem: hedging the risk of writing a European call option</td>
<td>106</td>
</tr>
<tr>
<td>3</td>
<td>The Black and Scholes option pricing model and delta hedging</td>
<td>107</td>
</tr>
<tr>
<td>3.1</td>
<td>Black and Scholes Option Pricing Model</td>
<td>108</td>
</tr>
<tr>
<td>3.2</td>
<td>The Leland Model: introduction of transaction costs</td>
<td>112</td>
</tr>
<tr>
<td>3.3</td>
<td>The delta of a call option and delta hedging</td>
<td>112</td>
</tr>
<tr>
<td>4</td>
<td>Minimax hedging strategy</td>
<td>114</td>
</tr>
<tr>
<td>4.1</td>
<td>Minimax problem formulation</td>
<td>114</td>
</tr>
<tr>
<td>4.2</td>
<td>The worst case scenario</td>
<td>115</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Worst Case 1</td>
<td>115</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Worst Case 2</td>
<td>115</td>
</tr>
<tr>
<td>4.3</td>
<td>The hedging error</td>
<td>116</td>
</tr>
<tr>
<td>4.4</td>
<td>The objective functions</td>
<td>117</td>
</tr>
<tr>
<td>4.4.1</td>
<td>The objective of minimax hedging</td>
<td>117</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Objective Function 1</td>
<td>117</td>
</tr>
<tr>
<td>4.4.2.1</td>
<td>Objective Function 1 defined</td>
<td>118</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Objective Function 2</td>
<td>121</td>
</tr>
<tr>
<td>4.4.3.1</td>
<td>Objective Function 2 defined</td>
<td>122</td>
</tr>
<tr>
<td>4.5</td>
<td>The minimax hedging error</td>
<td>124</td>
</tr>
</tbody>
</table>
4.6 Transaction Costs
  4.6.1 Transaction costs in the cost of hedging
  4.6.2 The transaction cost term in Objective Function 2
4.7 The variants of the minimax hedging strategy
4.8 The minimax solution

5 Simulation study of the performance of different strategies
  5.1 Objective of the simulation
  5.2 Generation of data
  5.3 Groups
  5.4 From set-up to wind-down
    5.4.1 Setting up the hedge
    5.4.2 Rebalancing the hedge
    5.4.3 Winding down the hedge
  5.5 Stratification
  5.6 Results of the simulation study
    5.6.1 Performance of minimax variants against delta hedging
      5.6.1.1 Main results from Table 5.6.1
    5.6.2 Relative performance between groups
      5.6.2.1 Main results from Table 5.6.2
    5.6.3 Relative performance between variants
      5.6.3.1 Main results from Table 5.6.3
  5.7 Statistical testing of hypotheses
    5.7.1 Hypotheses on the relative performance of the variants
      5.7.1.1 Comments on rejected hypotheses
    5.7.2 Hypotheses on the relative performance between groups
      5.7.2.1 Comments on rejected hypotheses
    5.7.3 Hypotheses on the relative performance between variants
      5.7.3.1 Comments on rejected hypotheses
  5.8 Intersigma performance of D*, E and F
    5.8.1 Main results from the analysis ... D*, E and F

6 Empirical study of the performance of different strategies
  6.1 The objective of the empirical study
  6.2 From set-up to wind-down
  6.3 Results of the empirical study
    6.3.1 Relative performance of variants, with intergroup comparisons
      6.3.1.1 Main results from Table 6.3.1a and Table 6.3.1b
    6.3.2 Relative performance between variants
      6.3.2.1 Main results from Table 6.3.2

page 5
8 CONCLUSION

1 The proposed quasi-Newton algorithm for continuous minimax
   1.1 Description of the proposed algorithm and numerical experiments
   1.2 Theoretical and numerical results
   1.3 Future directions

2 Applications of the proposed algorithm to risk management in finance
   2.1 The development of different hedging strategies
   2.2 Results from the simulation and empirical studies
   2.3 Future directions

3 Finale: the relation of continuous minimax to risk management

REFERENCES
# LIST OF TABLES

## 3 NUMERICAL RESULTS

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4</td>
<td>Published optimal values for Test Problems 16-20</td>
<td>95</td>
</tr>
</tbody>
</table>

## 4 MINIMAX HEDGING STRATEGY

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7</td>
<td>The hedging strategies to be used in the empirical and simulation studies</td>
<td>127</td>
</tr>
<tr>
<td>5.3a</td>
<td>The idealized groups</td>
<td>133</td>
</tr>
<tr>
<td>5.3b</td>
<td>The actual groups</td>
<td>135</td>
</tr>
<tr>
<td>5.6.1</td>
<td>Relative performance of variants by group and by sigma</td>
<td>139</td>
</tr>
<tr>
<td>5.6.2</td>
<td>Difference in relative performance between groups</td>
<td>141</td>
</tr>
<tr>
<td>5.6.3</td>
<td>Difference in relative performance between variants</td>
<td>143</td>
</tr>
<tr>
<td>5.7.1</td>
<td>Composite results of tests on the relative performance of the variants</td>
<td>145</td>
</tr>
<tr>
<td>5.7.2</td>
<td>Composite results of tests on the relative performance between groups</td>
<td>146</td>
</tr>
<tr>
<td>5.7.3</td>
<td>Composite results of tests on the relative performance between variants</td>
<td>148</td>
</tr>
<tr>
<td>5.8</td>
<td>Differences in relative performance between high performing variants</td>
<td>149</td>
</tr>
<tr>
<td>6.3.1a</td>
<td>Relative performance of the variants for 30 options</td>
<td>156</td>
</tr>
<tr>
<td>6.3.1b</td>
<td>Relative performance ... in the empirical and simulation study</td>
<td>156</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Difference in relative performance between selected pairs of variants</td>
<td>157</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Results of tests on relative performance and intergroup performance</td>
<td>158</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Results of tests on intervariant performance</td>
<td>159</td>
</tr>
</tbody>
</table>

## 5 MULTI-PERIOD MINIMAX HEDGING STRATEGIES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5a</td>
<td>Relative performance of three minimax strategies</td>
<td>175</td>
</tr>
<tr>
<td>4.5b</td>
<td>Difference in relative performance between minimax strategies</td>
<td>175</td>
</tr>
<tr>
<td>4.6</td>
<td>The accept/reject state of the hypotheses, for each level of sigma</td>
<td>178</td>
</tr>
<tr>
<td>4.7</td>
<td>Rank order of positive significant differences ... for each level of sigma</td>
<td>179</td>
</tr>
</tbody>
</table>

## 6 CAPM-BASED MINIMAX HEDGING STRATEGY

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>Stratification of the strategies according to the treatment of options</td>
<td>196</td>
</tr>
<tr>
<td>3.5a</td>
<td>Relative performance of strategies</td>
<td>197</td>
</tr>
<tr>
<td>3.5b</td>
<td>Difference in performance between minimax strategies</td>
<td>197</td>
</tr>
<tr>
<td>3.6</td>
<td>The accept/reject states of the hypotheses</td>
<td>199</td>
</tr>
<tr>
<td>3.7a</td>
<td>Rank order of positive significant differences ... for an individual option</td>
<td>200</td>
</tr>
<tr>
<td>3.7b</td>
<td>Rank order of positive significant differences ... for a portfolio of options</td>
<td>201</td>
</tr>
</tbody>
</table>

## 7 PERFORMANCE OF THE ALGORITHM AS A COMPUTATIONAL TOOL FOR THE HEDGING STRATEGIES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Performance when algorithms were applied to a portfolio of three options and when they were applied to the individual options</td>
<td>229</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

4  MINIMAX HEDGING STRATEGY
Figure 2.3  Profit or loss graph at the expiration date  106
Figure 5.8a The effect of sigma and moneyness on ... performance of variants  151
Figure 5.8b Variation with sigma of the high performing variants, E, F and D*  153
Figure A2 Relative performance of variants E and F ... K=0.02 and K=0.005  163

5  MULTI-PERIOD MINIMAX HEDGING STRATEGIES
Figure 5.1.1a. The stock price series of an option is deep-in-the-money  181
Figure 5.1.1b. The minimax hedging errors and actual hedging errors  181
Figure 5.1.2a. The stock price series of an option is to-at-the-money  182
Figure 5.1.2b. The minimax hedging errors and actual hedging errors  182
Figure 5.1.3a. The stock price series of an option is out-the-money  183
Figure 5.1.3b. The minimax hedging errors and actual hedging errors  183

6  CAPM-BASED MINIMAX HEDGING STRATEGY
Figure 4 Beta-risk profile using CAPM minimax  202
Introduction

This dissertation combines research in optimization, in particular continuous minimax, and in risk management, in particular the management of derivative securities. We outline below the main coverage of the chapters. We also describe the structure and format of the chapters.

In Chapter 1, we review the basic concepts in continuous minimax and its optimality conditions, present three continuous minimax algorithms, and discuss their limitations.

In Chapter 2, we present the proposed quasi-Newton algorithm that uses a quadratic approximation with a penalty term, justify the use of this penalty term, and present proofs of convergence.

In Chapter 3, we test the algorithm for different types of problems namely: convex-concave and convex-convex continuous minimax problems, and published discrete minimax problems. We report the solutions found by the proposed algorithm and the solutions found by another minimax algorithm and by a nonlinear programming algorithm, where appropriate.

In Chapter 4, we start the application to risk management in finance. We consider the problem of hedging the risk of writing options and develop a minimax hedging strategy to
deal with this risk. We propose several variants of the minimax hedging strategy and identify conditions under which they outperform a benchmark strategy. We test the performance of these variants against the benchmark strategy for different levels of volatility and different degrees of moneyness.

In Chapter 5, we extend the strategy developed in Chapter 4 to a multi-period setting. We consider a cautious hedger who wishes to avoid potentially large hedging errors for the second period in case he does not rebalance at the end of the first period; we develop a minimax strategy with a two-period setting. We also consider an aggressive hedger who wishes to monitor actual hedging error and rebalance when it becomes unacceptable to him; we develop a minimax strategy that includes a decision rule on early rebalancing.

In Chapter 6, we present a minimax strategy that deals with market movements, and apply it to the problem of hedging the risk of holding a portfolio of written call options.

In Chapter 7, we look at the performance of the proposed algorithm when it is applied to the risk management of options. This chapter is an extension of Chapter 3 where the performance of the proposed algorithm for a number of test problems was discussed.

In Chapter 8, we briefly summarize the major developments and the results, give conclusions and suggest extensions to the current work.

Each chapter is subdivided into sections. In our numbering system, we give the main section number followed by a dot and followed by the subsection number. For example, Section 2.3.4 means Subsubsection 4 in Subsection 3 in Section 2. Any reference to a section within a chapter is given by its section number. Any reference to a section from another chapter is given by the chapter number and section number, for example, Section 2.3.4 of Chapter 4.
Tables are local to the section and they are identified by their section number. When there are several tables in a section, alphabetic identification follows the table number. For example, respectively, Table 2.3.4a and Table 2.3.4b are the first and second tables in Section 2.3.4. Figures follow the same convention used for tables.

Equation identifications are local to the main section. These are numbered consecutively with the main section number as a localization index. For example, Eqn (2.2) means the second numbered equation in Section 2 of the chapter.

Footnotes are local to the chapter. These are numbered consecutively from 1. Appendices are also local to the chapter. A reference list that gives an alphabetical listing of all references cited in the chapters follows Chapter 8.
1 Introduction

In this chapter, we give the general background to continuous minimax and identify the foundation of an algorithm we develop to solve the continuous minimax problem. After presenting the problem and the corresponding optimality conditions, we discuss the limitations of algorithms designed to solve smooth problems when they are applied to the non-smooth problem of continuous minimax. We give three algorithms designed to solve the continuous minimax problem, and note their limitations. Finally, we discuss the relevance of global optimization algorithms in the context of continuous minimax for finding global maximizers as solutions to the maximization subproblem.

2 Basic concepts in continuous minimax

We consider the minimization of a real function $\Phi(x)$

$$\min_{x \in \mathbb{R}^n} \Phi(x)$$  \hspace{1cm} (2.1)

where

$$\Phi(x) = \max_{y \in Y} f(x, y)$$  \hspace{1cm} (2.2)

\textsuperscript{1}In Appendix 1, 2 and 3, respectively, we report the algorithms developed by Chaney[5], Panin[27] and Kiwiel[19].
and $Y$ is a bounded closed subset of $\mathbb{R}^m$. We call $\Phi(x)$ the max-function. We denote the set $Y(x)$ as the set of maximizers

$$Y(x) = \{y(x) \in Y \mid y(x) = \arg \max_{y \in Y} f(x, y)\}. \quad (2.3)$$

The function $f(x, y)$ is continuous in both $x$ and $y$, and continuously differentiable with respect to $x$ on $\mathbb{R}^n \times Y$. $\Phi(x)$ is then piecewise $C^1$, i.e. $\mathbb{R}^n$ is composed of regions inside which the gradient $\nabla \Phi(x)$ exists and is continuous, and at the boundary of which $\nabla \Phi(x)$ jumps (although $\Phi(x)$ itself is continuous). At this boundary, we have a kink, a point where $\nabla \Phi(x)$ is non-unique. $\Phi(x)$ has a gradient whenever $f(x, y)$ is maximized at a single $y$, and differentiability usually fails when $y$ is no longer unique. The subdifferential $^2$ of $\Phi(x)$ at $x$, denoted by $\partial \Phi(x)$, is given by

$$\partial \Phi(x) = \text{conv}\{g \mid g = \nabla_x \Phi(x) = \nabla_x f(x, y), y \in Y(x)\}. \quad (2.4)$$

where $\text{conv}\{\cdot\}$ denotes convex hull$^4$. The subdifferential is a non-empty convex compact set which reduces to the gradient in case $\Phi(x)$ has a unique derivative at $x$. The elements of $\partial \Phi(x)$ are called subgradients. The directional derivative$^5$, denoted by $\Phi'(x, d)$, is given by

$$\Phi'(x, d) = \lim_{t \to 0} \frac{1}{t} [\Phi(x + td) - \Phi(x)]. \quad (2.5)$$

The directional derivative is the support function$^6$ of $\partial \Phi(x)$, i.e.

$$\Phi'(x, d) = \max_{g \in \partial \Phi(x)} < g, d >. \quad (2.6)$$

---

$^2$Proposition 3H in Rockafellar[35, p35].
$^3$This has also been referred to as the Clarke[8] generalized gradient in Polak[31].
$^4$From Demyanov and Malozemov[11], the convex hull of an arbitrary set $G \subset \mathbb{R}^r$ is defined by

$$\text{conv}\{G\} = \left\{ X = \sum_{k=1}^r \alpha_k X_k \mid \alpha_k \geq 0; \sum_{k=1}^r \alpha_k = 1; r = 1, 2, \ldots \right\}$$

$^5$Theorem 2.1 in Demyanov and Malozemov[11, p188].
$^6$Definition 2.2.4 in Polak[30].
For a fixed $x$, $\Phi'(x,d)$ is convex in $d$. As such, it has a subdifferential which is $\partial \Phi(x)$.

3 The optimality condition

If $x^*$ is a local minimizer of Problem (2.1), then the following equivalent statements hold:

i) $\Phi(x^*,d) \geq 0$, $\forall d \in \mathbb{R}^n$;  \hspace{1cm} (3.1)

ii) $0 \in \partial \Phi(x^*)$. \hspace{1cm} (3.2)

Hence, the set $X^*$ of optimal points is characterized by

$$X^* = \{ x^* \in \mathbb{R}^n \mid 0 \in \partial \Phi(x^*) \}. \hspace{1cm} (3.3)$$

4 The differentiability of the max-function

The max-function is continuous but is, in general, non-differentiable in the sense of smooth functions; the presence of kinks accounts for this. Demyanov and Malozemov[11,Theorem 2.1, p188] state that the function $\Phi(x)$ is differentiable; this is not in the sense of smooth functions but in the sense of Clarke's[8] general differentiability. In line with Clarke's definition of differentiability, Rockafellar [35]\(^8\) states that $\Phi(x)$ is strictly differentiable at $x$ if $\partial \Phi(x)$ consists of a single vector, i.e. $\partial \Phi(x) = \{ \nabla \Phi(x) \}$.

5 The Hessian of the max-function

Wierzbicki[46] has established the existence of the Hessian of the max-function and its formulation for the discrete minimax problem. In Chapter 2, we extend the results from

---

\(^7\)Theorem 5.1 in Polak[31].

\(^8\)Theorem 4F, p43, in Rockafellar[35].
Wierzbicki [46] to the continuous case and show that, subject to the underlying assumptions for the proposed algorithm, the Hessian is well-defined.

6. Failure of smooth methods

In this section, we discuss the main causes of failure in directly applying, without modification, a smooth method to Problem (2.1).

1. Failure of convergence

In a classical smooth method, we replace a smooth function \( \Phi(x) \) by a linear or quadratic model to approximate the directional derivative

\[
<V_0(x), d> \quad [\equiv \Phi(x + d) - \Phi(x)]
\]

\[
<V_0(x), d> + \frac{1}{2} <d, \nabla^2 \Phi(x) d> \quad [\equiv \Phi(x + d) - \Phi(x)]
\]

and minimize these models. Minimization of (6.1) on the unit ball gives the steepest descent, and minimization of (6.2) gives Newton's method. For a non-smooth function such as (2.2), the above models are no longer defined at a kink and, close to a kink, they no longer provide an efficient approximation of \( \Phi(x) \). To approximate \( \Phi(x) \), (6.1) - (6.2) may need to be augmented or modified\(^10\).

2. Failure of optimality test

Another possible problem is the lack of an implementable stopping rule. For a \( C^1 \)-function, the gradient will become small in norm, say

\[
|V \Phi(x)| \leq \varepsilon, \quad \varepsilon > 0 \text{ small.}
\]

\(^9\)Gill, Murray and Wright [16] discuss the how these approximating functions are used in nonlinear programming.

\(^10\)Zang [47] developed a smoothing technique for discrete minimax optimization where an approximating function replaces the original function in the neighborhood of a kink.
When $x_k$ approaches some optimal $x_*$, this can be used to stop the algorithm. However, for non-smooth $\Phi(x)$, (6.3) is not a suitable a stopping rule, even if the gradient existed at all iterates. The reason for this is the presence of subgradients at a kink where (6.3) is not well defined. With condition (2.4), a stopping rule based on the optimality condition (3.2) implies an exploration of the whole $y$-space in order to fully define $Y(x_*)$.

3. Failure of gradient approximation

For nonsmooth functions, computing subgradients is compulsory and finite differencing may cause failure. In the case of smooth functions, the differential $\delta\Phi = \langle \nabla \Phi, \delta x \rangle$ varies linearly with the differential $\delta x$. In the nonsmooth case, $\delta\Phi$ is not linear and the differential must be estimated by

$$\delta\Phi = \frac{[\Phi(x + td) - \Phi(x)]}{t}$$

(6.4)

for all directions $d \in \mathbb{R}^n$.

The algorithm developed by Panin[27] to solve Problem (2.1) uses a modified quadratic approximation based on (2.6) and (6.2). Panin's modified quadratic approximation, together with Wierzbicki's[46] convex duality theory, was adapted by Kiwiel[19]. The assumptions underlying Chaney's[5] algorithm imply that the max-function is continuously differentiable and therefore amenable to the use of smooth methods. We propose an algorithm in Chapter 2 to solve Problem (2.1) that uses a modified quadratic approximation based on (2.6), (6.2), and on results from Wierzbicki[46]. All algorithms presented in this chapter, including the proposed algorithm presented in Chapter 2, solve Problem (2.1) by directly approximating (6.4); a modified finite-differencing method has not been attempted.
7 Methods to solve continuous minimax

Problems in continuous minimax are in the general class of non-smooth problems. As such, they may be solved using methods developed in non-smooth optimization, e.g. subgradient and bundle methods\textsuperscript{11}. Subgradient methods require at least one subgradient to be evaluated at each iteration in order to find a direction of descent. Bundle methods use subgradient information from successive iterations, within a ball of some radius. These methods were developed to solve either the general class or particular types of non-smooth problems. Those addressing the general class lack explicit steps to deal with the maximization subproblem, while those solving particular types, such as discrete minimax, are too particular to be applicable to continuous minimax\textsuperscript{12}.

In the following sections, we suggest the limitations of three algorithms (Chaney\textsuperscript{[5]}, Panin\textsuperscript{[27]} and Kiwiel\textsuperscript{[19]}) designed to solve Problem (2.1). We propose a quasi-Newton algorithm and present this in Chapter 2.

8 Comments on Chaney's algorithm\textsuperscript{13}

Chaney's\textsuperscript{[5, 1982]} algorithm is highly restrictive in practice because it requires numerous assumptions to be satisfied before it can be applied. His assumption 4 is particularly restrictive\textsuperscript{14}: it restricts the algorithm to problems where the max-function has no kinks. For problems with no constraints, the assumptions on concavity in the \(y\)-space and convexity in the \(x\)-space restrict the algorithm to saddle point problems. For the problems on which it can be used, this algorithm is well-specified and readily implementable.

\textsuperscript{11}A comprehensive review of non-smooth methods can be found in Lemarechal\textsuperscript{[21]}.\textsuperscript{12}Shor\textsuperscript{[45]} developed a non-smooth optimization algorithm that can be applied to discrete minimax problems.\textsuperscript{13}See Appendix 1.\textsuperscript{14}Kiwiel\textsuperscript{[19]} noted that Chaney's\textsuperscript{[5]} algorithm require a unique maximizer for all \(x\).
9 Comments on Panin's algorithm\textsuperscript{15} 

Panin\textsuperscript{[27, 1981]} uses a quadratic approximation similar to (6.2) for smooth functions. However, the quadratic approximation is maximized and this marks the departure from (5.2). His algorithm is not implementable: in Step 1 he does not specify how to solve the minimax problem with the quadratic approximation as objective function. It is assumed that one can find the minimizer of the equivalent max-function of the quadratic approximation. The interplay between the minimization subproblem and the maximization subproblem is not specified.

10 Comments on Kiwiel's algorithm\textsuperscript{16} 

Kiwiel\textsuperscript{[19, 1987]} adopted the quadratic approximation used by Panin\textsuperscript{[27]} to develop an algorithm to solve Problem 2.1. Kiwiel uses a direction-finding algorithm that attempts to find a direction of descent. Step 2 of the direction finding algorithm assumes that a maximizer of the linear approximation can be found. There is no restriction on the choice of maximizer. In contrast, in the proposed quasi-Newton algorithm (presented in Chapter 2), we restrict the choice of maximizer by using a penalty term. This ensures that the maximizer corresponds to the minimum norm subgradient and thereby leads to a direction of descent.

In Kiwiel's direction-finding algorithm, it is possible to find the same maximizer for the linear approximation for iteration $t-1$ and for iteration $t$. If, on using this maximizer, the optimality condition in Step 2 of the direction-finding algorithm is not satisfied, then $d_t$ is the same as $d_{t-1}$. The direction-finding algorithm loops infinitely without satisfying Kiwiel's Theorem 4.1(ii). This looping can be avoided by a careful choice of maximizer at the start of the direction-finding algorithm.

\textsuperscript{15}See Appendix 2.
\textsuperscript{16}See Appendix 3.
In Kiwiel's algorithm, the number of maximizations of the approximating function per iteration of the main algorithm is the number of iterations of the direction finding algorithm less one. In contrast, in the proposed quasi-Newton algorithm, the number of maximizations is fixed at one.

11 The relevance of global optimization algorithms

The continuous minimax problem given by (2.1) requires a solution to (2.2) in order to define the max-function. This is a maximization subproblem of which the solution is required to be a global maximizer in order for the max-function to be well-defined. Some minimax problems may have a structure in which the global maximizer may be identified using a nonlinear programming algorithm. Examples of such problems are those that are concave in the y-space, in particular, saddle point problems. Other minimax problems may have a structure where the global maximizer is always an extreme point, or where the global maximizer may itself be a saddle point solution. The importance of the use of global maximizers at each iteration of a minimax algorithm is that the algorithm can guarantee a monotonic decrease of the max-function.

Developments in the field of global optimization, particularly the development of algorithms to find global maximizers, are relevant to continuous minimax. A comprehensive survey of global algorithms can be found in Pardalos and Rosen[28] and in Rinnooy Kan and Timmer[34].
Appendix 1
Chaney's Algorithm

1 The Problem

Chaney[5] developed an algorithm for solving
\[ \min_{x \in X} \max_{y \in Y} f(x, y) \]  
such that
\[ X = \{ x \in \mathbb{R}^n : g_i(x) \leq 0, i = 1, \ldots, t \} \]  
\[ Y = \{ y \in \mathbb{R}^m : h_j(y) \leq 0, j = 1, \ldots, s \} \]

At each fixed \( x \),
\[ \Phi(x) = \max_{y \in Y} f(x, y) \]  
is called the inner problem
\[ \min_{x \in X} \Phi(x) \]  
is called the outer problem.

2 Notation

Let \( x \in X \), \( y \in \mathbb{R}^m \). The following quadratic programming problem, denoted by \( QP(x, y) \),
\[ \min D + \frac{1}{2} \| d \|^2 \]  
subject to
\[ D \in \mathbb{R}^l, \ d \in \mathbb{R}^n. \]
\[ < \nabla_x f(x, y), d > \leq D \]
\[ g_i(x) + < \nabla g_i(x), d > \leq D, \ i = 1, \ldots, t. \]

We shall denote the solution to Problem (A1.6) as \( D(x, y) \) and \( d(x, y) \).

At each iteration, there exists \( w = (w_0, w_1, \ldots, w_t) \in W_t \) in \( \mathbb{R}^{1+t} \), for which \( \sum_{i=0}^{t} w_i = 1 \) and \( w_i \geq 0 \), so that
\[ d + w_0 \nabla_x f(x, y(x)) + \sum_{i=1}^{t} w_i \nabla g_i(x) = 0 \]  
\[ w_0 [< \nabla \Phi(x), d > - D] = 0 \]
\[ w_i [g_i(x) + < \nabla g_i(x), d > - D] = 0, \ i = 1, \ldots, t. \]

This notation reduces to the Karush Kuhn-Tucker conditions for the solution \( x^* \) of the outer problem.

3 Assumptions
1. \( g_i(x) \leq 0, i = 1, \ldots, t \) are real-valued functions in \( \mathbb{R}^n \).
2. \( h_j(y) \leq 0, j = 1, \ldots, s \) are real-valued functions in \( \mathbb{R}^m \).
3. \( f(x, y) \), \( g_i(x) \leq 0, i = 1, \ldots, t \), \( h_j(y) \leq 0, j = 1, \ldots, s \), are all twice continuously differentiable functions.
4. For each \( x \in X \), the problem of maximizing \( f(x, y) \) over \( Y \) has a unique solution denoted by \( Y(x) \).
5. The sets \( X \) and \( Y \) are compact.
6. A number \( \delta > 0 \) exists so that, for each \( x \in X \), the set \( V(x) \) of all vectors \( (v_0, v_1, \ldots, v_s) \) for which \( v_0 \geq \delta, \ v_j \geq 0 \) for each \( j = 1, \ldots, s \).
\[ \sum_{j=0}^{s} v_j = 1, \quad (A1.8) \]

\[ v_j h_j(Y(x)) = 0, \quad \text{for } j \geq 1, \]

\[ -v_0 \nabla_y f(x, Y(x)) + \sum_{j=1}^{s} v_j \nabla_y h_j(Y(x)) = 0 \]

is non-empty. (It is assumed that the Karush-Kuhn-Tucker conditions hold at the solution \( Y(x) \) to the inner problem.)

7. A positive number \( m_1 \) exists so that

\[ \left( d, \left\{ -v_0 \nabla^2_{yy} f(x, Y(x)) + \sum_{j=1}^{s} v_j \nabla^2_{yy} h_j(Y(x)) \right\} d \right) > m_1 |d|^2 \]

whenever \( x \in X, \nu \in V(x) \) and \( d \) any non-zero feasible direction for which

\[ < \nabla_y h_j(Y(x)), d > = 0 \quad \text{for all } j \text{ in } \{ j: v_j > 0 \}. \]

8. Let \( M \geq 1 \) be a number and \( U^* \) and \( W^* \) be bounded neighborhoods of \( X \) and \( Y \) such that

\[ \| \nabla^2_{xx} f(x, y) \| \leq M, \quad \| \nabla^2_{xy} f(x, y) \| \leq M, \quad \| \nabla^2 g_i(x) \| \leq M \quad (A1.10) \]

whenever \( x \) is in \( U^* \), \( y \) is in \( W^* \) and \( i = 1, \ldots, t \).

9. Suppose the Pironneau-Polak[29] algorithm was used to generate an infinite sequence \( \{ x_k \} \) which converges to \( x^* \) where \( x^* \) solves the outer problem. Let \( \{ w_k \} \) be the sequence associated with \( \{ x_k \} \). Assume that \( s_0 > 0 \) where

\[ s_0 = \inf \{ w_0 : w \text{ is the limit point of } \{ w_k \} \}. \]

10. A positive number \( m_3 \leq 1 \) exists so that

\[ \left( d, \left\{ -\nabla^2_{yy} f(x, x) + \sum_{j=1}^{s} w_j(x) \nabla^2_{yy} h_j(Y(x)) \right\} d \right) > m_3 \]

whenever \( w \) is any limit point of \( \{ w_k \} \) and \( d \) is any feasible direction for which

\[ < \nabla g_i(x), d > = 0 \quad \text{whenever } i \text{ is such that } w_i > 0. \]

(This condition is a version of the standard second-order sufficiency condition; it clearly holds in case each \( g_i(x) \) is convex and for each fixed \( y \), \( f(x, y) \) is strongly convex.)

11. Assume that, given \( x \in X \), any \( y \in \mathbb{R}^m \), we have a globally convergent algorithm \( A(x, y) \) which can be used to generate a sequence \( \{ z \} \) convergent to \( Y(x) \) with \( z_1 = y \).

12. Assume that, given \( x \in X \), a test function \( t(x, \cdot) \) on \( \mathbb{R}^m \) is known so that positive numbers \( p_1, p_2, \) and \( \rho \geq 1 \) exist for which

\[ p_1 |y - Y(x)|^\rho \leq t(x, y) \leq p_2 |y - Y(x)|^\rho \quad (A1.12) \quad \text{whenever } x \in X \text{ and } y \in W^*_*, \]

where \( W^*_* \) is a bounded neighborhood of \( Y \).

---

17 Assumptions 9 and 10 are not relevant to the main algorithm but are used in establishing the convergence of an outer algorithm.

13. Assume that the test function $r(x, \bullet)$ on $\mathbb{R}^m$ is continuous on $X^* W$. (Hence, for $x \in X$ and a sequence $\{y_j\}_{j=1}^\infty$ in $W$, it follows that $\{y_j\}_{j=1}^\infty$ converges to $Y(x)$ if and only if the sequence $\{r(x,y_j)\}_{j=1}^\infty$ converges to 0.)

14. Assume that the algorithm $A(x, \bullet)$, $x \in X$, are not only globally convergent but that they are also "uniformly Q-quadratically convergent", that is, we shall assume that positive numbers $B$ and $\delta$ exist so that if $x \in X$, $|y - Y(x)| \leq \delta$ and $\tilde{y}$ is obtained after one iteration of $A(x,y)$, then $|\tilde{y} - Y(x)| \leq B|y - Y(x)|^2$. (A1.13)

15. Assume that for each $x$ in the neighborhood of $X$, it is true that $f(x, \bullet)$ and $h_j(y)$ are three-times continuously differentiable on $\mathbb{R}^m$.

16. Given $y \in \mathbb{R}^m$, put $I(y) = \{j: 1 \leq j \leq s, h_j(y) \geq 0\}$. We shall assume that the set $\{\nabla h_j(y); j \in I(y)\}$ is linearly independent for every $y \in \mathbb{R}^m$.

17. Assume that for each $x \in X$, $Y(x)$ is the only Kuhn-Tucker point for the inner problem. Thus, there are non-negative numbers $w_1(x), w_2(x), \ldots, w_s(x)$ so that $w_j(x)h_j(Y(x)) = 0$, for all $j = 1, \ldots, s$

and

$$-\nabla_y f(x, Y(x)) + \sum_{j=1}^s w_j \nabla h_j(Y(x)) = 0. \quad \text{(A1.14)}$$

If $v(x)$ is as in Assumption 6, we see that $w_j(x) = \frac{v_j(x)}{v_0(x)}$ for each $j = 1, \ldots, s$.

18. For each $x \in X$, assume that

$$\left\langle d, \left[ -\nabla^2_{yy} f(x, Y(x)) + \sum_{j=1}^s w_j(x) \nabla^2_{yy} h_j(Y(x)) \right] d \right\rangle > 0 \quad \text{(A1.15)}$$

whenever $d$ is any vector for which $< \nabla h_j(Y(x)), d > > 0$ for all $j$ for which $w_j(x) > 0$.

4 The Algorithm

Step 0: Initialization.

Choose the following parameters:

- $\beta \in (0,1)$,
- $\delta_0 \in (0,1)$,
- $\{\epsilon_t\}_{t=0}^\infty$ a strictly decreasing sequence convergent to zero,
- $\{\gamma_t\}_{t=0}^\infty$ a strictly decreasing sequence convergent to zero

with $\epsilon_t = o(\gamma_t)$,

- $z_{-1} \in W$,
- $x_0 \in X$,

and

Set $k = 0$.

19 The following assumptions are stricter versions of earlier assumptions which are needed to ensure that the test function used in the algorithm satisfies Assumptions 12 and 13.
Set \( l = 0 \),
Set \( z = z_{-1} \),
Set \( \mu_{-1} = 1 \).

**Step 1:** Apply Algorithm \( A(x_k, z) \) to obtain \( z_l \) so that
\[ |f(x_k, z_l)| \leq \epsilon_l. \]

**Step 2:** Solve problem \( QP(x_k, z_l) \) to get \( D(x_k, z_l) \) and \( d(x_k, z_l) \).
If \( D(x_k, z_l) + \frac{1}{2} |d(x_k, z_l)|^2 > -\gamma_l \),
Then:
\[ \begin{align*}
&\text{Set } z = z_l, \\
&\text{Set } \mu_{k+1} = \mu_l, \\
&\text{Set } l = l + 1, \\
&\text{Go to Step 1.}
\end{align*} \]
Otherwise:
\[ \begin{align*}
&\text{Set } z = z_l, \\
&\text{Set } y_k = z_l, \\
&\text{Go to Step 3.}
\end{align*} \]

**Step 3:** Perform the following steps.
(a) Set \( \alpha = 1 \).

(b) Apply Algorithm \( A(x_k + \alpha d(x_k, y_k), z) \) to get \( z^* \) so
\[ |f(x_k + \alpha d(x_k, y_k), z^*)| \leq \epsilon_l. \]
If \( g_i(x_k + \alpha d(x_k, y_k)) \leq \frac{1}{2} \alpha D(x_k, y_k) \) for each \( i = 1, \ldots, t \),
and If \( f(x_k + \alpha d(x_k, y_k), z^*) - f(x_k, y_k) \leq \frac{1}{2} \alpha D(x_k, y_k) \),
Then:
\[ \begin{align*}
&\text{Set } \alpha_k = \alpha, \\
&\text{Set } \mu_{l+1} = \mu_l, \\
&\text{Set } x_{k+1} = x_k + \alpha_k d(x_k, y_k), \\
&\text{Set } z = z^*, \\
&\text{Set } k = k + 1, \\
&\text{Set } l = l + 1, \\
&\text{Go to Step 1.}
\end{align*} \]
Otherwise:
\[ \begin{align*}
&\text{Go to Step 3(c).}
\end{align*} \]

(c) If \( \alpha \leq \mu_l \),
Then:
\[ \begin{align*}
&\text{Go to Step 3(d).}
\end{align*} \]
Otherwise:
\[ \begin{align*}
&\text{Set } z = z^*, \\
&\text{Set } \alpha = \beta \alpha, \\
&\text{Go to Step 3(b).}
\end{align*} \]

(d) Set \( z = z_l \),
Set \( \mu_{l+1} = \delta_0 \mu_l \),
Set \( \alpha_k = 0 \),
Set \( x_{k+1} = x_k \),
Set \( k = k + 1 \).
Set \( l = l + 1 \).
Go to Step 1.

5 Description of the algorithm
1. The purpose of \( E_l = o(\gamma_l) \) is to help force the solution of the inner problem to be more rapid than that of the outer problem.

2. Step 1 is designed to give, at any stage, an improved solution to the current inner problem.

3. In Step 2, a decision is made as to whether we can try to improve our current 'solution' \( x_k \) to the outer problem. We proceed to Step 3 only when the test in Step 2 suggests that we are now much closer to a solution \( Y(x_k) \) of the current inner problem than we are to a solution of the outer problem.

4. In Step 3, we perform a line search of the Armijo type from \( x_k \) in the direction \( d(x_k, y_k) \). For each prospective value \( x_k + \alpha d(x_k, y_k) \), we find it necessary to search for a point \( z^* \) which is suitably close to \( Y(x_k + \alpha d(x_k, y_k)) \).

In case we arrive at Step 3(d), the line search is abandoned as a failure. Chaney proved that this failure can occur only a finite number of times.

6 On the construction of a test function which satisfies Assumptions 12-13
This test function is to serve as a monitoring device for the process of solving the inner problem of finding \( Y(x) \). The test function is defined in terms of a certain augmented Lagrangian associated with the inner problem.

Definition of terms in the test function:
Let \( x \in X \), and \( y \in \mathbb{R}^m \). Define \( \lambda(x, y) = (\lambda_1, \ldots, \lambda_s) \) in \( \mathbb{R}^s \) to be the minimizer of the expression

\[
\left| -\nabla_y f(x, y) + \sum_{j=1}^s \lambda_j \nabla h_j(y) \right|^2 + \sum_{j=1}^s \lambda_j^2 h_j(y)^2. \tag{A1.16}
\]

Let \( c > 0 \) and let \( a(x, y, c) \) in \( \mathbb{R}^m \) be:

\[
a_j(x, y, c) = h_j(y) \quad \text{if} \quad c h_j(y) + \lambda_j \geq 0 \tag{A1.17}
\]

\[
a_j(x, y, c) = -\frac{\lambda_j}{c} \quad \text{if} \quad c h_j(y) + \lambda_j < 0 \tag{A1.18}
\]

Let the augmented Lagrangian \( F \) be

\[
F(x, y, c) = -f(x, y) + \frac{1}{2c} \sum_{j=1}^s \left[ \max(0, c h_j(y) + \lambda_j) \right]^2 - \lambda_j^2 \tag{A1.19}
\]

Finally, let the test function be:

\[
t(x, y, c) = -|\nabla_y F(x, y, c)|^2 + \frac{|a(x, y, c)|^2}{c}. \tag{A1.20}
\]

This test function is used as part of Step 1 and Step 3(b).
Appendix 2
Panin’s Algorithm

1 The Problem
Panin[27] developed an algorithm to solve

\[
\min_{x \in X} \max_{y \in Y} f(x, y)
\]

(A2.1)

2 Notation
Let the quadratic approximation to the max-function at each iteration be given by \( f_k(d, y) \) defined as

\[
f_k(d, y) = f(x_k, y) + \langle \nabla_x f(x_k, y), d \rangle + \frac{a}{2} d^2
\]

(A2.2)

where \( a > 0 \) is a constant. Eqn (A2.2) is a linearization of \( f(x, y) \) at \( x_k \) plus a term that restricts the norm of \( d \). This is a convex function, and as such, it has a solution.

Define \( \Phi_k(d) \) as

\[
\Phi_k(d) = \max_{y \in Y} f_k(d, y).
\]

(A2.3)

Problem (A2.1) is solved by considering the problem

\[
\min_{d \in X} \Phi_k(d).
\]

(A2.4)

Let

\[
Y(x) = \{ y(x) \in Y | y(x) = \arg \max_{y \in Y} f(x, y) \},
\]

(A2.5)

\[
Y_{k+1} = \{ y_{k+1} \in Y | y_{k+1} = \arg \max_{y \in Y} f_k(d, y) \}
\]

(A2.6)

be subsets of \( Y \) and

\[
\Psi_k = \min_{d \in X} \Phi_k(d) - \Phi(x_k)
\]

(A2.7)

approximate the directional derivative.

3 Assumptions
1. \( X \subset \mathbb{R}^n \) and \( Y \subset \mathbb{R}^m \) are convex compact sets.
2. \( f(x, y) \) and \( \nabla_x f(x, y) \) are continuous.
3. For any \( y \in Y \) and any \( x_1, x_2 \in \mathbb{R}^n \), we have

\[
|\nabla_x f(x_1, y) - \nabla_x f(x_2, y)| \leq L|x_1 - x_2|
\]

(A2.8)

where \( L > 0 \) is a constant and \(|\cdot|\) is the norm in \( \mathbb{R}^n \).
4. \( \nabla^2_{xx} f(x, y) \) is continuous in the neighborhood of the unique point \( \{x_*, y_*\} \) where \( x_* \in X \) and \( y_* \in Y(x_*) \).
5. There exist numbers \( M \geq m > 0 \) such that for any \( x \in \mathbb{R}^n \)

\[
mlx^2 \leq \langle x, \nabla^2_{xx} f(x_*, y_*)x \rangle \leq Mlx^2.
\]

(A2.9)

4 The Algorithm
Step 0: Initialization.
Select \( x_1 \).
Set \( k = 1 \).

Step 1:
Solve for \( \Phi(x_k) \)
\[ \Phi(x_k) = \max_{y \in Y} f(x_k, y), \]
Solve \( \min_{d \in X} \Phi_k(d) \).
Set \( \Psi_k = \min_{d \in X} \Phi_k(d) - \Phi(x_k) \).

Step 2:
Set \( \alpha = 1 \),
Set \( \varepsilon = 1 \),
(a) If \( \Phi(x_k + \alpha d_k) - \Phi(x_k) \leq \varepsilon \alpha \Psi_k \),
Then:
Set \( x_{k+1} = x_k + \alpha d_k \),
Go to Step 3.
Otherwise:
Set \( \alpha = \frac{\alpha}{2} \),
Set \( \varepsilon = \frac{\varepsilon}{2} \),
Go to Step 2(a).

Step 3:
If \( \alpha_k < 1 \),
Then:
Set \( a = 2a \).

5 Description of the Algorithm
In Step 1, we solve (A2.4) with the assumption that we can solve (A2.3). The maximizer of the quadratic approximation is assumed to correspond to the minimizer found in Step 1. This is the weakness of the algorithm: it is not clear how one can solve (A2.4).
Appendix 3
Kiwiell's Algorithm

1 The Problem
Kiwiell[19] developed an algorithm to solve
\[ \min_x \max_y f(x,y) \]  
(A3.1)

2 Notation
Let the quadratic approximation to the max-function at each iteration be given by \( f_k(d,y) \) defined as
\[ f_k(d,y) = f(x_k, y) + \langle \nabla_x f(x_k, y), d \rangle + \frac{1}{2} \| d \|^2 \]  
(A3.2)

Eqn (A3.2) is a linearization of \( f(x, y) \) at \( x_k \) plus a term that restricts the norm of \( d \). This is a convex function, and as such, it has a solution.

Define \( \Phi_k(d) \) as
\[ \Phi_k(d) = \max_{y \in Y} f_k(d,y) \]  
(A3.3)

Problem (A3.1) is solved by considering the problem
\[ \min_{d \in X} \Phi_k(d) \]  
(A3.4)

Let
\[ Y(x) = \{ y(x) \in Y | y(x) = \arg \max_{y \in Y} f(x, y) \} \]  
(A3.5)
\[ Y_{k+1} = \{ y_{k+1} \in Y | y_{k+1} = \arg \max_{y \in Y} f_k(d,y) \} \]  
(A3.6)

be subsets of \( Y \) and
\[ \Psi_k = \min_{d \in X} \Phi_k(d) - \Phi(x_k) \]  
(A3.7)

approximate the directional derivative.

Finally, for any \( x \in X \) and \( \varepsilon > 0 \), let \( \Phi_{\varepsilon}(x) \) be any number satisfying
\[ \Phi(x) - \varepsilon \leq \Phi_{\varepsilon}(x) \leq \Phi(x) \]  
(A3.8)

3 Assumptions
1. \( Y \subset \mathbb{R}^m \) is compact.
2. \( f(x, y) \) and \( \nabla_x f(x, y) \) are both continuous in \( \mathbb{R}^n \).
3. For any \( d \in \mathbb{R}^n \) and \( \varepsilon > 0 \) one can find a point \( y \in Y \) such that
\[ f(x,y) + \langle \nabla_x f(x,y), d \rangle \geq \Phi_k(d) - \varepsilon \]

4 The Algorithm
Step 0: Initialization.
Select \( x_1, y_1 \),
Select final accuracy \( \varepsilon \geq 0 \),
Select linesearch parameter \( c, 0 < c < 1 \),
Select stepsize parameter \( \lambda, 0 < \lambda < 1 \),
Select a linear approximation parameter \( m, 0 < m < 1 \),

---

This quadratic approximation is similar to (A2.2) except for the coefficient of the last term.
Set \( k = 1 \).

**Step 1:** Maximization at current point \( x_k \).
Solve \( \Phi(x_k) = \max_{y \in \mathcal{Y}} \{ f(x_k, y) \} \). (A3.9)

**Step 2:** Direction-finding subproblem.
Set \( x = x_k \) and use the direction-finding algorithm with the parameters \( \varepsilon, \gamma \geq 0 \) and \( m \) until it stops, returning \( d_k \) and \( \Psi_k \).
If \( \Psi_k \geq -\varepsilon \), then terminate.

**Step 3:** Line search.
Find \( \alpha_k \) from
\[
\alpha_k = \max\{ \alpha : \Phi(x_k + \alpha d) - \Phi(x_k) \leq c\alpha \Psi_k, \alpha = (\lambda)^i, i = 0, 1, 2, \ldots \} \] (A3.10)
Set:
\[
x_{k+1} = x_k + \alpha_k d_k, \\
k = k + 1,
\]
Go to Step 1.

5 Direction-Finding Algorithm (Step 2 of Kiwiel's Algorithm)
Data: \( x_k \in \mathbb{R}^n, \Phi(x_k), \varepsilon, \gamma \geq 0, m \in (0, 1) \).

**Step 0:** Initialization.
Set \( x = x_k \) and \( \Phi(x_k) = \Phi(x) \).
Select any \( y \in \mathcal{Y} \) and set \( p_0 = \nabla_x f(x, y) \), \( \Theta_0 = f(x, y) \).
Set \( t = 1 \).

**Step 1:** Reduced dual solving.
Find a number \( \mu_t \) to solve
\[
\min_{\mu \in \mathbb{R}} \left\{ \frac{1}{2} \left| \left( 1 - \mu \right) p_{t-1} + \mu \nabla_x f(x, y) \right|^2 - (1 - \mu) \Theta_{t-1} - \mu f(x, y) \right\} \] (A3.12)
Set \( p_t = (1 - \mu_t) p_{t-1} + (\mu_t) \nabla_x f(x, y) \)
(A3.13)
Set \( \Theta_t = (1 - \mu_t) \Theta_{t-1} + (\mu_t) f(x, y) \)
(A3.14)
Set \( \Psi_t = -(|p_t|^2 + \Phi(x) - \Theta_t) \)
(A3.15)
If \( \Psi_t \geq -\varepsilon \), then terminate returning \( d_k = -p_t \) and \( \Psi_k = \Psi_t \).
(A3.16)
Otherwise, proceed.

**Step 2:** Primal optimality testing.
Set \( d_t = -p_t \).
Find
\[
y_{t+1} = \arg \max_{y \in \mathcal{Y}} \{ f(x, y) + \langle \nabla_x f(x, y), d_t \rangle \} \] (A3.17)
If \( f(x, y_{t+1}) + \langle \nabla_x f(x, y_{t+1}), d_t \rangle > -\Phi(x) \leq m \Psi_t \)
then terminate returning \( d_k = -p_t \) and \( \Psi_k = \Psi_t \).
Otherwise set \( t = t + 1 \) and go to step 1.

6 Description of the algorithm
The direction-finding algorithm attempts to find the minimum norm gradient that is an element of the subdifferential of the approximating function, and uses this gradient in defining the descent direction. The
approximation to the directional derivative given by (A3.15) is based on Lemma 1 of Wierzbicki[46]. Kiwiel revised the algorithm by using inexact evaluations. This involves the use of Assumption 3. The revised algorithm assumes that a finite process can find δ-solutions to the maximization problem

$$\max_{y \in Y} f_k(d, y).$$

Inexact line searches and a mechanism for decreasing δ ensure global convergence of the method to stationary points of $\Phi(x)$, i.e. points satisfying the optimality condition of the original problem. The use of inexact evaluations of $\Phi(x)$ at line searches ensure convergence without spending too much effort at initial iterations, when low accuracy suffices.

\[21 \text{ In Chapter 3, we present Kiwiel's algorithm with exact evaluations and note that the implementations of both Kiwiel's and the proposed quasi-Newton algorithm make use of another algorithm to solve the maximization subproblem with a predefined } \delta. \text{ We presented Kiwiel's algorithm with exact evaluations to make the presentation of all algorithms consistent.} \]
1 Introduction

In this chapter, we develop an algorithm for the continuous minimax problem; the proposed algorithm uses a quasi-Newton search direction conditional on the approximate maximizer. In Section 1.1, we give a general background to the problem that we wish to solve. In Section 2, we introduce the definitions of the main functions used in the algorithm. In Section 3, we present the algorithm. In Section 4, we justify the search direction subproblem. In Section 5, we discuss basic convergence results leading to the monotonic decrease of the sequence \( \{ \Phi(x_t) \} \). In Section 6, we discuss convergence to unit stepsizes, global and local convergence results.

1.1 The continuous minimax problem

We wish to solve the problem

\[
\min_{x \in \mathbb{R}^n} \max_{y \in Y} f(x, y).
\]

In (1.1) we are given a compact infinite set \( Y \subset \mathbb{R}^m \) and a continuous function \( f: \mathbb{R}^n \times Y \to \mathbb{R} \) such that the gradient \( \nabla_x f(x, y) \) with respect to \( x \) is continuous on \( \mathbb{R}^n \times Y \).
Let

\[ \Phi(x) = \max_{y \in Y} f(x, y) \]  

(1.2)

for all \( x \in \mathbb{R}^n \). We call \( \Phi(x) \) the \textbf{max-function}. We note that (1.2) has a solution since \( f(x, y) \) is continuous on the compact set \( Y^2 \). The continuous minimax problem is to solve

\[
\min_{x \in \mathbb{R}^n} \Phi(x).
\]

(1.3)

If \( Y \) is a finite set, (1.3) is known as the discrete minimax problem. Problem (1.1) in the discrete case is given by

\[
\min_{x \in \mathbb{R}^n} \max_{i} f_i(x)
\]

(1.4)

This problem is widely studied and a number of algorithms\(^3\) have been proposed to solve (1.4). Most of these algorithms involve the transformation of (1.4) into a constrained nonlinear programming problem as

\[
\min_{x, z} \left\{ z \mid f_j(x) \leq z, j = 1, \ldots, i \right\}
\]

(1.5)

This method for solving the discrete minimax problem cannot be used for the continuous minimax case. Formulation (1.5) for the continuous case is

\[
\min_{x, z} \left\{ z \mid f(x, y) \leq z, \forall y \in Y \right\}
\]

(1.6)

This is a problem in semi-infinite optimization where we have an infinite number of constraints corresponding to the infinite number of elements in the set \( Y \).

\(^2\)This is Weierstrass' Theorem, in Luenberger[22].

\(^3\)See, for example, Rustem[40], Dutta and Vidyasagar[13], Charalambous and Conn[6].
Problem (1.3) poses several difficulties. First, as discussed in Chapter 1, the function \( \Phi(x) \) is in general continuous but may not be straightforwardly differentiable. This is because \( \Phi(x) \) may have kinks. The presence of kinks makes the optimization problem difficult to solve. At a kink, the maximizer is not unique; therefore the choice of subgradient to use in finding a search direction is not a simple task compared to smooth functions. This further implies that the Hessian of \( \Phi(x) \), if it exists, has to be viewed in the context of multiple maximizers. Wierzbicki proved the existence of the Hessian of \( \Phi(x) \) when the set \( Y \) is finite, i.e. the max-function of discrete minimax. Second, \( \Phi(x) \) may not be computed accurately because this would require infinitely many iterations of an algorithm for maximizing \( f(x,y) \) over \( Y \). In practice, the algorithm for solving the maximization problem has to be stopped when a sufficiently good maximum value has been attained. Third, problem (1.3) requires a global maximum in view of possible multiple maximizers such as corner solutions. The use of non-global maxima cannot guarantee a monotonic decrease in \( \Phi(x) \).

We develop a quasi-Newton minimization method that defines a descent direction conditional on the approximate maximizer. The latter is determined using an augmented maximization to ensure that multiplicity of maximizers does not result in an inferior search direction. The algorithm takes a step along this search direction. In the maximization subproblem, we restrict the choice of maximizer to that which corresponds to the minimum norm subgradient (Lemma 4). This ensures a direction of descent (Lemma 2). The stepsize is regulated by a merit function.

---

4This practical problem does not affect the theory. Also in practical applications discussed in Chapter 3, the potential effects of this difficulty have not been observed.

5This too is a practical problem. In applications considered in Chapter 3, special precautions are taken to check for extreme point solutions (corner solutions for upper and lower bounded problems) as well as possible (local) mid-range solutions that would normally be observed in convex-concave problems.
2 Definitions

Let \( f_k(d,y) \) denote the quadratic approximation to \( f(x,y) \)

\[
f_k(d,y) = f(x_k,y) + \langle \nabla_x f(x_k,y), d \rangle + \frac{1}{2} |d|_{H_k}^2
\]  
(2.1)

and

\[
\Phi_k(d) = \max_{y \in Y} f_k(d,y)
\]  
(2.2)

where \( |d|_{H_k}^2 = \langle d, H_k d \rangle \) and \( H_k \) is a positive definite Hessian approximation to the Hessian with respect to \( x \) of \( \Phi(x) \) at the \( k^{th} \) iteration. The reason for this choice of Hessian and not using an approximation to \( \nabla_x^2 f(x,y) \) is that, by a convex duality theory given as Lemma 1 in Wierzbicki[46], the problem of finding the minimum norm subgradient in the subdifferential of the max-function is equivalent to the problem of minimizing the quadratic approximation to the Lagrangian of (1.6) which is (2.2). The optimal choice of Hessian, therefore, is an approximation to the Hessian of the Lagrangian of (1.6).\(^7\)

Let

\[
Y(x) = \left\{ y \in Y \mid y = \arg \max_{y \in Y} f(x,y) \right\}.
\]  
(2.3)

We can rewrite the initial problem as

\[
\min_{x \in \mathbb{R}^n} \Phi(x).
\]  
(2.4)

\(^6\)The existence of the Hessian of \( \Phi(x) \) is established in the Appendix.

\(^7\)The equivalence between the Hessian of the Lagrangian of (1.6) and that of \( \Phi(x) \) is also established in the Appendix.
To solve (2.4), we construct the iterative sequence \( x_{k+1} = x_k + \alpha_k d_k \), where \( \alpha_k \) is calculated according to a rule discussed below while \( d_k \) is the solution of the auxiliary problem

\[
\min_{d \in \mathbb{R}^n} \Phi_k(d). \tag{2.5}
\]

Corresponding to the solution \( d_k \) of (2.5), we have the set of maximizers given by

\[
Y_{k+1} = \left\{ y_{k+1} \in Y \mid y_{k+1} = \arg \max_{y \in Y} f_k(d, y) \right\}. \tag{2.6}
\]

Finally, we define the approximate directional derivative as

\[
\Psi_k = \min_{d \in \mathbb{R}^n} \Phi_k(d) - \Phi(x_k). \tag{2.7}
\]

### 2.1 Assumptions

1. \( Y \subset \mathbb{R}^m \) is convex and compact.
2. \( f(x, y) \) is twice continuously differentiable both in \( x \) and \( y \).
3. There exists a number \( m > 0 \) such that \( m|x|^2 \leq |x|_H^2 \) for all \( x \in \mathbb{R}^n \).
4. In the neighborhood of the solution \( x^* \) of (2.4), there exists \( b > 0 \) such that
   \[ \forall x \in \{ x \mid |x - x^*| < b, x \in \mathbb{R}^n \} \text{ the Hessian with respect to } x \text{ of } f(x, y) \text{ for all } y \in Y \text{ is positive definite.} \]
5. \( \Phi(x) \) is bounded from below in \( \mathbb{R}^n \).

### 2.2 Necessary Condition for an Extremum (nce)

\[
\max_{y \in \mathcal{Y}(x^*)} \left( \nabla_x f(x^*, y), d \right) \geq 0, \quad \forall d \in \mathbb{R}^n \tag{2.8}
\]

\(^8\)This is required for demonstrating convergence to unit stepsize and local superlinear convergence.
3 The Algorithm

Step 0. Initialization.

Select $x_i, y_i, H_i$.

Select final accuracy $\xi \geq 0$ ($\xi = 1.0e^{-6}$).

Select linesearch parameter $c$, $0 < c < 1$, ($c = 1.0e^{-4}$).

Select stepsize parameter $\lambda$, ($\lambda = 5.0e^{-1}$).

Select penalty function coefficient $C$, $0 << C < \infty$, ($C = 1.0e^6$).

Set $k=1$.

Step 1. Maximization at current point $x_k$.

Solve $\Phi(x_k) = \max_{y \in Y} \{f(x_k, y)\}$. (3.1)

Step 2. Direction-finding subproblem.

Find $y_{k+1}$ from

$$y_{k+1} = \arg \max_{y \in Y} \{f(x_k, y) - \frac{1}{2} \| \nabla_x f(x_k, y) \|_{H_k}^2 - C(\Phi(x_k) - f(x_k, y))\} \tag{3.2}$$

Set:

$$d_k = -H_k^{-1} \nabla_x f(x_k, y_{k+1}) \tag{3.3}$$

$$\Psi_k = f(x_k, y_{k+1}) - \frac{1}{2} \| \nabla_x f(x_k, y_{k+1}) \|_{H_k}^2 - \Phi(x_k) \tag{3.4}$$

An alternative formulation of Eqn(3.2) is

$$y_{k+1} = \arg \max_{y \in Y} \left\{ f(x_k) - \frac{1}{2} \| \nabla_x f(x_k, y) \|_{H_k}^2 + C(\Phi(x_k) - f(x_k, y))^2 \right\} \tag{3.2a}$$

where the third term of (3.2) is squared in (3.2a).

The use of (3.2a) instead of (3.2) gives an alternative algorithm with the same properties as the original algorithm. All the proofs apply, with all references to the penalty term $C(\Phi(x_k) - f(x_k, y))$ changed to $C(\Phi(x_k) - f(x_k, y))^2$.

Numerical results discussed in Chapter 3 show that both alternatives have similar results.
If $\Psi_k \geq -\xi$, then terminate. \hfill (3.5)

Step 3. Line search.

Find $\alpha_k$ from
$$\alpha_k = \max \{ \alpha \mid \Phi(x_k + \alpha d_k) - \Phi(x_k) \leq c\alpha \Psi_k, \alpha = (\lambda)^i, i = 0, 1, 2, \ldots \} \hfill (3.6)$$

Set:
$$x_{k+1} = x_k + \alpha_k d_k$$
$$k = k + 1$$

Update the Hessian.

Go to Step 1.

4 Justification for the Search Direction Subproblem

As we move from $x$ to $x + d$, the change in objective function value $\Phi(x + d) - \Phi(x)$ can be approximated by $\Phi_k(d) - \Phi(x)$. Thus, we could find a search direction for $\Phi(x)$ by solving (2.5). In the algorithm, the Newton direction

$$d_k = -H_k^{-1}\nabla_x f(x_k, y_{k+1}) \hfill (4.1)$$

is used in finding the minimizer $d_k$ of $\Phi_k(d)$, where $y_{k+1}$ is the solution to the penalty function\(^7\)

$$y_{k+1} = \arg \max_{y \in Y} \left\{ f(x_k, y) - \frac{1}{2}||\nabla_x f(x_k, y)||^2_{H_k^{-1}} - C(\Phi(x_k) - f(x_k, y)) \right\} \hfill (4.2)$$

\(^7\)In (4.2) we can have multiple maximizers $y \in Y(x_k)$ solving $\Phi(x_k) = \max \{ f(x_k, y) \}$. Some of these $y$ values $y \in Y(x_k)$ do not solve (4.4) whereas others do. Numerical experiments indicate that the choice of $y \in Y(x_k)$ in (4.2) is quite important.
In justifying the use of this penalty function, we first consider the nonlinear programming problem
\[
\min_{x} f(x). \tag{4.3}
\]
This can be solved by minimizing a quadratic approximation to \( f(x) \) as
\[
\min_{d} f_k(d) \tag{4.4}
\]
where
\[
f_k(d) = f(x_k) + \langle \nabla_x f(x_k), d \rangle + \frac{1}{2} d^2 H_k. \tag{4.5}
\]

The first term of (4.5) is \( f(x_k) \) which implies that \( f_k(d) = f(x_k) \) when \( d = 0 \). In other words, the graph of \( f_k(d) \) touches the graph of \( f(x) \) at \( x_k \).

We consider the justification for the search direction for the minimax problem. Problem (1.1) is solved by considering the sequential quadratic approximation (2.1) and solving (2.2).

From Kiwiel[19, p 274], the change in objective function value \( \Phi(x + d) - \Phi(x) \), as we move from \( x \) to \( x + d \), is approximated by \( \Phi_k(d) - \Phi(x) \). When \( d = 0 \), (2.2) is reduced to
\[
\Phi_k(0) = \max_{y \in Y} \{ f(x_k, y) + 0 + 0 \} = \Phi(x_k). \tag{4.6}
\]
In other words, the graph of \( \Phi_k(d) \) touches the graph of the max function \( \Phi(x) \) at \( x_k \).

If \( d \neq 0 \), we need to find a \( \overline{y} \in Y \) such that the first term of (4.6) is as close to the max-function as possible to ensure that the quadratic approximation closely approximates the max-function, i.e.
\[
\Phi_k(d) = f(x_k, y) + \langle \nabla_x f(x_k, y), d \rangle + \frac{1}{2}d_H^2 \tag{4.7}
\]

where

\[
0 \leq \Phi(x) - f(x, y) \leq \delta, \quad \delta \geq 0 \text{ is a small number.} \tag{4.8}
\]

To compute \( y_{k+1} \), we use the Newton direction and reformulate (2.2) as

\[
\Phi_k(d) = \max_{y \in Y} \{ f(x_k, y) - \| \nabla_x f(x_k, y) \|^2_{H_k^{-1}} + \frac{1}{2} \| \nabla_x f(x_k, y) \|^2_{H_k^{-1}} \}. \tag{4.9}
\]

However, we also determine \( y \) to minimize \( \{ \Phi(x) - f(x, y) \} \). This ensures that the quadratic approximation is a good approximation to the max-function.

For a given value of \( x \), there may be a set of maximizers for (1.1). Let \( Y(x) \) denote this set. For a given \( x_k \), we solve the maximizing problem and denote its value as \( \Phi(x_k) \). We wish to restrict the solution of (4.9) to be as close to the value solving \( \Phi(x_k) \). In order to achieve this, we augment (4.9) with a penalty term. Consider, therefore, the solution

\[
y_{k+1} = \arg\max_{y \in Y} \{ f(x_k, y) - \frac{1}{2} \| \nabla f(x_k, y) \|^2_{H_k^{-1}} - C(\Phi(x_k) - f(x_k, y)) \} \tag{4.10}
\]

We note that, by the definition of \( \Phi(x_k) \), \( \Phi(x_k) - f(x_k, y_{k+1}) \geq 0 \). For some \( C \), \( 0 < C < \infty \), the solution of (4.9) is restricted by

\[
0 \leq \Phi(x_k) - f(x_k, y_{k+1}) \leq \delta \tag{4.11}
\]

where \( \delta \) is a small nonnegative number for a large \( C \).
We consider two scenarios:

Case 1: \[ y_{k+1} \Rightarrow \Phi(x_k) - f(x_k, y_{k+1}) = 0 \] (4.12)
Case 2: \[ y_{k+1} \Rightarrow 0 < \Phi(x_k) - f(x_k, y_{k+1}) \leq \delta \, , \text{ with small } \delta > 0 \] (4.13)

We now wish to justify that \( d \) determined by (4.1) is a direction of descent for \( y_{k+1} \) solving (4.9) for both these cases.

To discuss Cases 1 and 2, we need the definition of \( \Psi_k \) which is used in the stepsize strategy below:

\[ \Psi_k = \min_d \Phi_k(d) - \Phi(x_k) \] (4.14)
\[ \Psi_k = f(x_k, y_{k+1}) - \frac{1}{2} \left| \nabla_x f(x_k, y_{k+1}) \right|_{H_k}^2 - \Phi(x_k). \] (4.15)

Case 1: \[ y_{k+1} \Rightarrow \Phi(x_k) - f(x_k, y_{k+1}) = 0 \] (4.16)

Thus, using (4.15), the solution \( y_{k+1} \) implies that

\[ \Psi_k = -\frac{1}{2} \left| \nabla_x f(x_k, y_{k+1}) \right|_{H_k}^2. \] (4.17)

Case 2: \[ y_{k+1} \Rightarrow 0 < \Phi(x_k) - f(x_k, y_{k+1}) = \delta_0 \, , \, 0 < \delta_0 \leq \delta \] (4.18)

Thus, using (4.15), the solution \( y_{k+1} \) implies that

\[ \Psi_k = -\frac{1}{2} \left| \nabla_x f(x_k, y_{k+1}) \right|_{H_k}^2 - \delta_0 \] (4.19)
\[ \Psi_k < -\frac{1}{2} \left| \nabla_x f(x_k, y_{k+1}) \right|_{H_k}^2. \] (4.20)

This shows that \( d \) is a direction of descent for both cases.
5 Basic Convergence Results

Lemma 1. Let Assumptions 1 and 2 hold. Condition $d_k = 0$ is necessary and sufficient for point $x_k$ to satisfy the nce of the initial problem.

Proof.

Problem (2.5) (minimize the convex approximation), can be expressed as

$$
\min_{d \in \mathbb{R}^n} \Phi_k(d) = \min_{x \in \mathbb{R}^n} \max_{y \in Y_{k+1}} \{ f(x_k, y) + \langle \nabla_x f(x_k, y), d \rangle + \frac{1}{2} \| d \|^2_{H_k} \}. \quad (5.1)
$$

Let $d_k$ be the minimizer of $\Phi_k(d)$. Then (5.1) becomes

$$
\Phi_k(d_k) = \max_{y \in Y_{k+1}} \{ f(x_k, y) + \langle \nabla_x f(x_k, y), d_k \rangle + \frac{1}{2} \| d_k \|^2_{H_k} \}. \quad (5.2)
$$

Let $d_k = 0$. Then (5.2) becomes

$$
\Phi_k(0) = \max_{y \in Y_{k+1}} \{ f(x_k, y) \}. \quad (5.3)
$$

The nce of (5.2) when $d_k = 0$ is

$$
\frac{\partial \Phi_k(d)}{\partial (d)} = \max_{y \in Y_{k+1}} \{ \nabla_x f(x_k, y), d \} \geq 0 \quad (5.4)
$$

Since $f_k(d, y) = f(x_k, y)$, this implies that $Y_{k+1} = Y(x_k)$. Then (5.4) becomes

$$
\frac{\partial \Phi_k(d)}{\partial (d)} = \max_{y \in Y(x_k)} \{ \nabla_x f(x_k, y), d \} \geq 0 \quad (5.5)
$$

which is the nce of the original problem.
Conversely, let the nce be satisfied at point $x_k = x_*$.

Denote

$$f_*(x - x_*, y) = f(x_*, y) + \langle \nabla_x f(x_*, y), x - x_* \rangle + \frac{1}{2} \|x - x_*\|^2_{H_k} \quad (5.6)$$

$$\Phi_*(x - x_*) = \max_{y \in Y} \{ f_*(x - x_*, y) \} \quad (5.7)$$

$$\min_{x \in \mathbb{R}^n} \Phi_*(x - x_*) \quad (5.8)$$

Let $\bar{x}$ be the solution of (5.8) and let $y_* = \arg \max f_*(\bar{x}, y)$. Problem (5.7) becomes

$$\Phi_*(\bar{x} - x_*) = \max_{y \in Y} \{ f(x_*, y) + \langle \nabla_x f(x_*, y), \bar{x} - x_* \rangle + \frac{1}{2} \|\bar{x} - x_*\|^2_{H_*} \}. \quad (5.9)$$

For the nce of (5.8), we have

$$\max_{y \in Y(\bar{x})} \{ \langle \nabla_x f(x_*, y) + H_*(\bar{x} - x_*), \bar{x} - x_* \rangle \geq 0 \}. \quad (5.10)$$

Since the nce of the original problem is satisfied if $\bar{x} = x_*$, and as we have let $x_k = x_*$, we have

$$Y(\bar{x}) = \{ y \in Y \mid y = \arg \max f_*(\bar{x}, y) \} = Y(x_*) = Y(x_k) \quad (5.11)$$

and hence $d_k = 0$.

QED.
Consider the function
\[
\Psi_k = \min_{d \in \mathbb{R}^n} \Phi_k(d) - \Phi(x_k)
\]  
and let \( d_k \) be the solution of (2.5). Problem (5.12) becomes
\[
\Psi_k = \max_{y \in Y_{k+1}} \{ f(x_k, y) + \langle \nabla_x f(x_k, y), d_k \rangle + \frac{1}{2} \| d_k \|_{H_k}^2 \} - \Phi(x_k)
\]  
or
\[
\Psi_k = f(x_k, y_{k+1}) + \langle \nabla_x f(x_k, y_{k+1}), d_k \rangle + \frac{1}{2} \| d_k \|_{H_k}^2 - \Phi(x_k)
\]  

Lemma 2. Descent Direction \( d_k \)

Let Assumptions 1 and 2 hold. For \( d_k \) computed by the algorithm we have
\[
\Psi_k \leq -\frac{1}{2} \| d_k \|_{H_k}^2 \leq 0.
\]  
Furthermore, if Assumption 3 holds, then
\[
\Psi_k \leq -\frac{1}{2} m \| d_k \|^2 \leq 0.
\]  

Proof.

From Section 3,
\[
0 \leq \Phi(x_k) - f(x_k, y_{k+1}) \leq \delta
\]  
where \( \delta \geq 0 \) is a small number. This implies that
\[
f(x_k, y_{k+1}) - \Phi(x_k) \leq 0.
\]
Using the direction $d_k = -H_k^{-1}\nabla_x f(x_k, y_{k+1})$, (5.14) becomes

$$
\Psi_k = f(x_k, y_{k+1}) - \left| \nabla_x f(x_k, y_{k+1}) \right|_{H_k^{-1}}^2 + \frac{1}{2} \left| \nabla_x f(x_k, y_{k+1}) \right|_{H_k^{-1}}^2 - \Phi(x_k) \quad (5.19)
$$

or

$$
\Psi_k = f(x_k, y_{k+1}) - \frac{1}{2} \left| \nabla_x f(x_k, y_{k+1}) \right|_{H_k^{-1}}^2 - \Phi(x_k). \quad (5.20)
$$

In view of (5.18) and the relation $\nabla_x f(x_k, y_{k+1}) = -H_k d_k$, (5.20) becomes

$$
\Psi_k \leq -\frac{1}{2} \left| d_k \right|_{H_k}^2 \leq -\frac{1}{2} m \left| d_k \right|^2 \leq 0. \quad (5.21)
$$

QED.
Corollary 2.1. Condition $\Psi_k = 0$ is necessary and sufficient for the nce of the original problem to be satisfied at $x_k$.

Proof.

Let $\Psi_k = 0$. This implies that

$$\min_{d \in \mathbb{R}^n} \Phi_k(d) = \Phi(x_k). \tag{5.22}$$

Let $d_k$ be the minimizer of $\Phi_k(d)$. Then

$$\Phi_k(d_k) = \Phi(x_k) = \Phi_k(0). \tag{5.23}$$

This implies that $d_k = 0$. From Lemma 1, this further implies that the nce of the original problem is satisfied at $x_k$.

Conversely:

Let $d_k = 0$. Then

$$\Psi_k = \Phi_k(d_k) - \Phi(x_k) = \Phi_k(0) - \Phi(x_k) = \Phi_k(0) - \Phi_k(0) = 0. \tag{5.24}$$

QED.
Lemma 3. Let Assumptions 1, 2 and 3 hold. Condition $d_k \to 0$ is satisfied if and only if $\Psi_1 \to 0$.

Proof.

Let $d_k \to 0$.

$$
\Phi_k(d_k) = \max_{y \in H_{k+1}} \{ f(x_k, y) + \langle \nabla_x f(x_k, y), d_k \rangle + \frac{1}{2}|d_k|^2 \} 
$$

(5.25)

$$
g \geq \max_{y \in Y(x_k)} \{ f(x_k, y) + \langle \nabla_x f(x_k, y), d_k \rangle + \frac{1}{2}|d_k|^2 \} 
$$

(5.26)

since $y \in Y(x_k)$ is not necessarily the maximizer at $x_k + d_k$.

$$
\Phi_k(d_k) \geq \Phi(x_k) + \max_{y \in Y(x_k)} \{ \langle \nabla_x f(x_k, y), d_k \rangle + \frac{1}{2}|d_k|^2 \} 
$$

(5.27)

or

$$
\Phi_k(d_k) - \Phi(x_k) \geq \max_{y \in Y(x_k)} \{ \langle \nabla_x f(x_k, y), d_k \rangle + \frac{1}{2}|d_k|^2 \} 
$$

(5.28)

The equation for $\Psi_k$ can now be expressed as

$$
\Psi_k = \Phi_k(d_k) - \Phi(x_k) \geq \max_{y \in Y(x_k)} \{ \langle \nabla_x f(x_k, y), d_k \rangle + \frac{1}{2}|d_k|^2 \} 
$$

(5.29)

If $d_k \to 0$, then $\Phi_k(d_k) \to \Phi(x_k)$. This implies that $\Psi_k \to 0$.

Conversely, let $\Psi_k \to 0$. Then from Lemma 2, $\Psi_k \leq -\frac{1}{2}m|d_k|^2 \leq 0$. Hence, $\Psi_k \to 0$ implies that $d_k \to 0$.

QED.
Lemma 4. Let Assumptions 1 and 2 hold and let $C$ be $0 < C < \infty$. From the algorithm,

$$y_{k+1} = \arg \max_{y \in \mathcal{F}} \left \{ f(x_k, y) - \frac{1}{2} \left| \nabla_x f(x_k, y) \right|_{H_k^{-1}}^2 - C(\Phi(x_k) - f(x_k, y)) \right \}$$  \hspace{1cm} (5.30)

implies that $y_{k+1} \in Y(x_k)$ and that

$$y_{k+1} = \arg \min_{y \in Y(x_k)} \left \{ \frac{1}{2} \left| \nabla_x f(x_k, y) \right|_{H_k^{-1}}^2 \right \}.$$  \hspace{1cm} (5.31)

Proof.

For a given $C$, $y_{k+1} \in Y \setminus Y(x_k)$ in (5.32) implies that

$$\Phi(x_k) - f(x_k, y_{k+1}) > 0.$$  \hspace{1cm} (5.32)

We can, by appropriately increasing $C \in (0, \infty)$, recompute $y_{k+1} \in Y$ to ensure that

$$\Phi(x_k) - f(x_k, y_{k+1}) = 0$$  \hspace{1cm} (5.33)

and consequently $y_{k+1} \in Y(x_k)$.

Therefore from the algorithm,

$$y_{k+1} = \arg \max_{y \in \mathcal{F}} \left \{ f(x_k, y) - \frac{1}{2} \left| \nabla_x f(x_k, y) \right|_{H_k^{-1}}^2 - C(\Phi(x_k) - f(x_k, y)) \right \}.$$  \hspace{1cm} (5.34)

implies that $y_{k+1} \in Y(x_k)$.

To prove that $y_{k+1} = \arg \min_{y \in Y(x_k)} \left \{ \frac{1}{2} \left| \nabla_x f(x_k, y) \right|_{H_k^{-1}}^2 \right \}$, it is the case that

$$f(x_k, y) = \Phi(x_k), \quad \forall y \in Y(x_k).$$  \hspace{1cm} (5.35)
Therefore,

\[
y_{k+1} = \arg \min_{y \in \Gamma(x_k)} \left\{ -f(x_k, y) + \frac{1}{2} \left| \nabla_x f(x_k, y) \right|^2_{H_k^{-1}} + C(\Phi(x_k) - f(x_k, y)) \right\} \tag{5.36}
\]

\[
y_{k+1} = \arg \min_{y \in \Gamma(x_k)} \left\{ -\Phi(x_k) + \frac{1}{2} \left| \nabla_x f(x_k, y) \right|^2_{H_k^{-1}} + C(\Phi(x_k) - \Phi(x_k)) \right\} \tag{5.37}
\]

\[
y_{k+1} = \arg \min_{y \in \Gamma(x_k)} \left\{ -\Phi(x_k) + \frac{1}{2} \left| \nabla_x f(x_k, y) \right|^2_{H_k^{-1}} + 0 \right\} \tag{5.38}
\]

Because the first and third terms do not affect the minimization problem,

\[
y_{k+1} = \arg \min_{y \in \Gamma(x_k)} \left\{ \frac{1}{2} \left| \nabla_x f(x_k, y) \right|^2_{H_k^{-1}} \right\}. \tag{5.39}
\]

This implies that

\[
y_{k+1} = \arg \max_{y \in \Gamma(x_k)} \left\{ f(x_k, y) - \frac{1}{2} \left| \nabla_x f(x_k, y) \right|^2_{H_k^{-1}} - C(\Phi(x_k) - f(x_k, y)) \right\}
\]

\[
= \arg \min_{y \in \Gamma(x_k)} \left\{ \frac{1}{2} \left| \nabla_x f(x_k, y) \right|^2_{H_k^{-1}} \right\}. \tag{5.40}
\]

QED.
Lemma 5. Let Assumptions 1 to 4 hold. For the strategy for computing $\alpha_k$,

$$\Phi(x_{k+1}) - \Phi(x_k) \leq c\alpha_k \Psi'$$  \hspace{1cm} (5.41)

the corresponding sequence of $\Phi(x_k), k=1,2,...,$ is monotonically decreasing.

Proof.

Second order expansion of $\Phi$

$$\Phi(x_k + \alpha_k d_k) = \Phi(x_k) + \alpha_k \max_{\nabla f(x_k,y) \in \partial \Phi(x_k)} < \nabla_x f(x_k,y), d_k >$$

$$+ \alpha_k^2 \int_0^1 (1-t)d_k \{ H[x_k + t(\alpha_k d_k)] \} dt > 0$$  \hspace{1cm} (5.42)

where $H[x_k + t(\alpha_k d_k)]$ is the Hessian of $\Phi(x)$ at a point between $(x_k)$ and $(x_k + \alpha_k d_k)$.

For every value of $x$, the Hessian exists since for $Y(x)$ a singleton, the Hessian is the same as that of $f(x,y)$ and for $Y(x)$ having more than one element, the Hessian is a linear combination of Hessians of $f(x,y), y \in Y(x)$. See Wierzbicki[46].

Because each of the elements of $Y(x_k)$ correspond to a subgradient in $\partial \Phi(x_k)$, we have the condition $y \in Y(x_k) \Rightarrow \nabla f(x_k,y) \in \partial \Phi(x_k)$. Thus, (5.42) can be expressed as

$$\Phi(x_k + \alpha_k d_k) = \Phi(x_k) + \alpha_k \max_{y \in Y(x_k)} < \nabla_x f(x_k,y), d_k >$$

$$+ \alpha_k^2 \int_0^1 (1-t)d_k \{ H[x_k + t(\alpha_k d_k)] \} dt > 0$$  \hspace{1cm} (5.43)

From Lemma 4, we have

$$y_{k+1} = \arg \min_{y \in Y(x_k)} |\nabla_x f(x_k,y)|^2_{H^{-1}}$$

$$= \arg \min_{y \in Y(x_k)} < \nabla_x f(x_k,y), H^{-1}_k \nabla_x f(x_k,y) >$$

$$= \arg \max_{y \in Y(x_k)} < \nabla_x f(x_k,y), d >.$$
For the second term of (5.43), (5.46) implies that the maximum is achieved at \( y_{k+1} \in Y(x_k) \).
Thus, (5.43) becomes

\[
\Phi(x_k + \alpha_k d_k) = \Phi(x_k) + \alpha_k \langle \nabla_x f(x_k, y_{k+1}), d_k \rangle \\
+ \alpha_k^2 \int_0^1 (1-t) \langle d_k, [H[x_t + t(\alpha_k d_k)]]d_k \rangle \, dt.
\] (5.47)

Without any loss of generality, \( \Phi(x_k) \) is obtained for any \( y \in Y(x_k) \), including \( y_{k+1} \).
Therefore, (5.48) is equivalent to

\[
\Phi(x_k + \alpha_k d_k) = f(x_k, y_{k+1}) + \alpha_k \langle \nabla_x f(x_k, y_{k+1}), d_k \rangle \\
+ \alpha_k^2 \int_0^1 (1-t) \langle d_k, [H[x_t + t(\alpha_k d_k)]]d_k \rangle \, dt
\] (5.48)

which can be expressed as

\[
\Phi(x_k + \alpha_k d_k) = f(x_k, y_{k+1}) + \alpha_k \langle \nabla_x f(x_k, y_{k+1}), d_k \rangle \\
+ \frac{1}{2} \alpha_k^2 \langle d_k, H_k \rangle
\] (5.49)

where \( H_k \) is a positive definite Hessian approximation to the Hessian with respect to \( x \) of \( \Phi(x) \) at the \( k^{th} \) iteration. The first three terms on the RHS are equal to \( \Phi_k(\alpha_k d_k) \), so (5.49) becomes

\[
\Phi(x_k + \alpha_k d_k) = \Phi_k(\alpha_k d_k) + \alpha_k^2 \rho_k \langle d_k \rangle^2
\] (5.50)

where

\[
\rho_k = \int_0^1 (1-t) \langle H[x_t + t(\alpha_k d_k)], H_k \rangle \, dt.
\] (5.51)

Since \( \Phi_k(d_k) \) is a convex function, we can write the first term on the RHS of (5.50) as

\[
\Phi_k(\alpha_k d_k) \leq \alpha_k \Phi_k(d_k) + (1-\alpha_k) \Phi_k(0)
\] (5.52)

\[
= \alpha_k \Phi_k(d_k) + \Phi_k(0) - \alpha_k \Phi_k(0)
\]

\[
= \Phi(x_k) + \alpha_k (\Phi_k(d_k) - \Phi(x_k))
\]

\[
= \Phi(x_k) + \alpha_k \Psi_k
\]

or

\[
\Phi_k(\alpha_k d_k) \leq \Phi(x_k) + \alpha_k \Psi_k.
\] (5.53)
Going back to (5.50),
\[ \Phi(x_k + \alpha_k d_k) \leq \Phi(x_k) + \alpha_k \Psi_k + \alpha_k^2 \rho_k |d_k|^2. \] (5.54)

From Lemma 2, \( \Psi_k \leq -\frac{1}{2} m |d_k|^2 \leq 0. \) (5.55)

Inequality (5.54) can be written as
\[ \Phi(x_k + \alpha_k d_k) - \Phi(x_k) \leq \alpha_k \Psi_k [1 + \frac{\alpha_k \rho_k}{\Psi_k} |d_k|^2]. \] (5.56)

In view of Inequality (5.55),
\[ \Phi(x_k + \alpha_k d_k) - \Phi(x_k) \leq \alpha_k \Psi_k [1 - \frac{2\alpha_k \rho_k}{m}]. \] (5.57)

For a given \( \rho_k \), there exists an \( \alpha_k \), \( 0 < \alpha_k \leq 1 \) such that
\[ 0 < c \leq 1 - \frac{2\alpha_k \rho_k}{m} < 1. \] (5.58)

Hence, given \( c \), there exists an \( \alpha_k \) such that Inequality (5.58) is satisfied. Since \( \Psi_k \leq -\frac{1}{2} m |d_k|^2 \leq 0 \), it further implies that \( \Phi(x_k) \) is monotonically decreasing.

QED.

Remark.

Since we are using an approximate Hessian in the expression for \( \rho_k \), we assume that
\[ \rho_k < P < \infty \quad , \text{where} \quad P > 0. \] (5.59)

52
Theorem 1 (Global Convergence): Let Assumptions 1 to 5 hold. The algorithm terminates at $x^*$ or it generates an infinite sequence in which there exists a subsequence \{x_k\} with $k \in K \subset \{1, 2, \ldots\}$ such that $\{d_k\} \to 0$ and every accumulation point $x'$ of the infinite sequence \{x_k\} is stationary for $\Phi(x)$.

Proof.

We first show that
\[
\lim_{k \to \infty} \left| \Psi_k \right| = 0
\]
(6.1)

From Lemma 5,
\[
\Phi(\alpha_k d_k) - \Phi(x_k) \leq c \alpha_k \Psi_k
\]
where $c \in (0, 1)$ and $0 < \alpha_k \leq 1$. From Lemma 2, $\Psi_k \leq 0$. Because $\Phi(x)$ is bounded on $\mathbb{R}^n$, $\Psi_k \leq 0$ implies that
\[
0 \leq c \sum_{k} \alpha_k |\Psi_k| \\
\leq \sum_k [\Phi(x_k) - \Phi(x_{k+1})] \\
< \infty
\]
(6.2)

which yields Condition (6.1).

To prove global convergence, let $x'$ be an accumulation point of \{x_k\}. Let there be a set $K \subset \{1, 2, \ldots\}$ such that $\{x_k\} \to x'$. Suppose that Condition (6.1) holds. Then, since $x_k \to x'$, we see that $x'$ is stationary because by Lemma 3, $\Psi_k \to 0 \Leftrightarrow d_k \to 0$ and by Lemma 1, $d_k = 0$ implies that $x_k = x' = x^*$ satisfies the nce.

QED.
**Theorem 2 (Unit Stepsize):** Let Assumptions 1 to 4 hold. Further, assume that the true Hessian is continuous in the neighborhood of the unique point \((x^*, y^*)\) where \(x^* \in X\) and \(y^* \in Y(x^*)\).

Let \(\{x_k\}\) be defined by the following:

\[
d_k = -H_k^{-1}\nabla_x f(x_k, y_{k+1})
\]
(6.3)

where

\[
y_{k+1} = \arg \max_{y \in Y} \left\{ f(x_k, y) - \frac{1}{2} |\nabla_x f(x_k, y)|_{H_k^{-1}}^2 - C(\Phi(x_k) - f(x_k, y)) \right\}
\]
(6.4)

and

\[
x_{k+1} = x_k + \alpha_k d_k, \quad 0 < \alpha_k \leq 1.
\]
(6.5)

Let the approximate Hessian \(H\) be positive definite with

\[
m|x|^2 \leq \langle x, Hx \rangle \leq M|x|^2, \quad 0 < m < M, \quad \forall x \in X.
\]
(6.6)

Let \(\alpha_1, 0 < \alpha_1 \leq 1\), be the largest number selected for which

\[
\Phi(x_{k+1}) - \Phi(x_k) \leq c\alpha_1 \Psi_k, \quad 0 < c < 1.
\]
(6.7)

Let \(\{x_k\}\) converge to \(x^*\). There exist two positive numbers \(t_1\) and \(t_2\), and integer \(K_0 > 0\) such that if

\[
|x_{k_0} - x^*| \leq t_1
\]
(6.8)

and

\[
|H_k - H^*| \leq t_2, \quad k \geq K_0
\]
(6.9)

Then \(\alpha_k = 1\).
(6.10)
Proof.
The second order expansion of $\Phi$ is given by (5.42) which has been expressed as (5.49).
This can also be expressed as

$$
\Phi(x_k + \alpha_k d_k) = f(x_k, y_{k+1}) + \alpha_k < \nabla_x f(x_k, y_{k+1}), d_k > + \frac{1}{2} \alpha_k^2 |d_k|^2 H_k \\
+ \alpha_k^2 \int_0^1 (1-t) < d_t, (H[x_t + t(\alpha_t d_t)] - H^* + H^* - H) d_t > dt
$$

(6.11)

where $H_k$ is a positive definite Hessian approximation to the Hessian with respect to $x$ of $\Phi(x)$ at the $k^{th}$ iteration and $H^*$ is the Hessian at the solution.

The first three terms on the RHS are equal to $\Phi_k(\alpha_k d_k)$, so (6.11) becomes

$$
\Phi(x_k + \alpha_k d_k) = \Phi_k(\alpha_k d_k) + \alpha_k^2 \rho_k |d_k|^2 + \alpha_k^2 \int_0^1 (1-t) < d_t, (H^* - H_k) d_t > dt
$$

(6.12)

where

$$
\rho_k = \int_0^1 (1-t) [H[x_t + t(\alpha_t d_t)] - H^*] dt.
$$

(6.13)

Applying the convexity property (as in (5.53))

$$
\Phi_k(\alpha_k d_k) \leq \Phi(x_k) + \alpha_k \Psi_k
$$

(6.14)

(6.12) becomes

$$
\Phi(x_k + \alpha_k d_k) \leq \Phi(x_k) + \alpha_k \Psi_k + \alpha_k^2 \rho_k |d_k|^2 + \alpha_k^2 \int_0^1 (1-t) < d_t, (H^* - H_k) d_t > dt
$$

(6.15)

and thus

$$
\Phi(x_k + \alpha_k d_k) \leq \Phi(x_k) + \alpha_k \Psi_k + \alpha_k^2 \rho_k |d_k|^2 + \alpha_k^2 < d_k, (H^* - H_k) d_k >.
$$

(6.16)

$$
\Phi(x_k + \alpha_k d_k) - \Phi(x_k) \leq \alpha_k \Psi_k [1 + \frac{\alpha_k \rho_k}{\Psi_k} |d_k|^2 + \frac{\alpha_k^2}{\Psi_k} < d_k, (H^* - H_k) d_k >]
$$

(6.17)
From Lemma 2, we have $\Psi \leq -\frac{1}{2} m|\bar{d}|^2 \leq 0$. Thus, (6.17) can be written as

$$
\Phi(x_k + \alpha_k \bar{d}_k) - \Phi(x_k) \leq \alpha_k \Psi_k \left[1 - \frac{2\alpha_k}{m} - \frac{2\alpha_k}{m} \left(\frac{1}{2} |H^* - H_k|\right)\right]
$$

(6.18)

$$
\Phi(x_k + \alpha_k \bar{d}_k) - \Phi(x_k) \leq \alpha_k \Psi_k \left[1 - \frac{2\alpha_k}{m} \left(\rho_k + \frac{1}{2} |H^* - H_k|\right)\right].
$$

(6.19)

From Lemma 5, there exists an $\alpha_k$, $0 < \alpha_k \leq 1$ such that

$$
0 < \alpha_k \leq 1 - \frac{2\alpha_k}{m} \left[\rho_k + \frac{1}{2} |H^* - H_k|\right] < 1.
$$

(6.20)

As \{x_k\} approaches $x^*$, by continuity, $\rho_k \to 0$ as $k \to \infty$. Therefore, for some radius $t_1$, $k > K_0$, and $|x_k - x^*| \leq t_1$ such that

$$
\rho_k < \frac{\alpha_k}{2}
$$

(6.21)

the inequality

$$
\rho_k + \frac{1}{2} |H^* - H_k| \leq \frac{\alpha_k}{2}
$$

(6.22)

holds provided that

$$
|H^* - H_k| \leq t_2 = 2(\frac{\alpha_k}{2} - \rho_k)
$$

(6.23)

The steplength $\alpha_k$ can be set equal to 1 provided that Inequality (6.23) is satisfied.

QED.
Theorem 3 (Q-superlinear convergence): Let Assumptions 1 to 4 hold. Further, let all the preconditions for Theorem 2 hold.

If
\[ \lim_{k \to \infty} \frac{|(H_k - H^*)(x_{k+1} - x_k)|}{|(x_{k+1} - x_k)|} = 0 \]  

(6.24)

then \( \{x_k\} \) is Q-superlinearly convergent.

Proof.

From Theorem 2, for
\[ |x_{K_0} - x^*| \leq t_1 \]  

(6.25)

and
\[ |H^* - H_k| \leq t_2, \quad k > K_0, \]  

(6.26)

the steplength \( \alpha_k \) can be set equal to 1. Therefore \( x_{k+1} = x_k + d_k \).

Consider the quadratic approximation \( \Phi_k(d) \)
\[ \Phi_k(d_k) \geq f_k(d_k, y_k) \]  

(6.27)

due to the fact that \( y_k \in Y \) is not necessarily a maximizer of \( f_k(d_k, y) \). Subtracting \( f(x_k, y_{k+1}) \) from both sides and invoking the value of (2.1), (2.2), and Lemma 4, we have

\[ \Phi_k(d_k) - f(x_k, y_{k+1}) \geq f_k(d_k, y_k) - f(x_k, y_{k+1}) \]  

(6.28)

\[ < \nabla_x f(x_k, y_{k+1}), d_k > + \frac{1}{2} |d_k|_{H_k}^2 \geq f_k(d_k, y_k) - f(x_k, y_{k+1}) . \]  

(6.29)

Using \( \nabla_x f(x_k, y_{k+1}) = -H_k d_k \), the LHS can be expressed as

\[ -\frac{1}{2} |d_k|_{H_k}^2 \geq f_k(d_k, y_k) - f(x_k, y_{k+1}) . \]  

(6.30)
The RHS can be expanded as

\[-\frac{1}{2} |d_k|_{H_k}^2 \geq f(x_k, y_k) + <\nabla_x f(x_k, y_k), d_k> + \frac{1}{2} |d_k|_{H_k}^2 - f(x_k, y_{k+1}). \tag{6.31}\]

Expanding each of the terms on the RHS of (6.31) using the quadratic approximation (2.1) with the true Hessian instead of \(H_k\), we have

First Term:

\[
f(x_k, y_k) = f(x_{k-1}, y_k) + <\nabla_x f(x_{k-1}, y_k), d_{k-1}> \\
+ \int_0^1 (1-t) <d_{k-1}, H[x_{k-1} + t(d_{k-1})]d_{k-1}> dt \\
= f_{k-1}(d_{k-1}, y_k) \\
+ \int_0^1 (1-t) <d_{k-1}, \{H[x_{k-1} + t(d_{k-1})] - H^*\}d_{k-1}> dt \\
+ \frac{1}{2} <d_{k-1}, (H^* - H_{k-1})d_{k-1} > \tag{6.32}
\]

Second Term:

\[
<\nabla_x f(x_k, y_k), d_k> = <\nabla_x f(x_{k-1}, y_k), d_k> \\
+ \int_0^1 <d_{k-1}, H[x_{k-1} + t(d_{k-1})]d_{k-1}> dt \\
= <\nabla_x f(x_{k-1}, y_k), d_k> + <d_{k-1}, H_{k-1}d_k> \\
+ \int_0^1 <d_{k-1}, \{H[x_{k-1} + t(d_{k-1})] - H^*\}d_k> dt \\
+ <d_{k-1}, (H^* - H_{k-1})d_k > \tag{6.33}
\]

Third Term:

\[
\frac{1}{2} <d_k, H_k d_k> \tag{6.34}
\]
Fourth Term:
\[ f(x_k, y_{k+1}) = f(x_{k-1}, y_{k+1}) + \nabla_x f(x_{k-1}, y_{k+1}) \cdot d_{k-1} + \int_0^1 (1-t) \nabla_x f(x_{k-1} + t(d_{k-1})) \cdot d_{k-1} \, dt \]
\[ = f_{k-1}(d_{k-1}, y_{k+1}) + \int_0^1 (1-t) \nabla_x f(x_{k-1} + t(d_{k-1})) \cdot d_{k-1} \, dt \]
\[ + \frac{1}{2} < d_{k-1}, (H^* - H_{k-1}) d_{k-1} > \]  \( (6.35) \)

Returning to (6.31), we have
\[ -\frac{1}{2} |d_k|^2_{H_k} \geq f(x_k, y_k) + \nabla_x f(x_k, y_k) \cdot d_k + \frac{1}{2} |d_k|^2_{H_k} - f(x_k, y_{k+1}) \]  \( (6.36) \)

\[ -\frac{1}{2} |d_k|^2_{H_k} \geq f_{k-1}(d_{k-1}, y_k) + \int_0^1 (1-t) \nabla_x f(x_{k-1} + t(d_{k-1})) \cdot d_{k-1} \, dt \]
\[ + \frac{1}{2} < d_{k-1}, (H^* - H_{k-1}) d_{k-1} > \]
\[ + < \nabla_x f(x_{k-1}, y_k), d_k > + < d_{k-1}, H_{k-1} d_k > \]
\[ + \int_0^1 < d_{k-1}, (H[x_{k-1} + t(d_{k-1})] - H^*) \cdot d_k > \, dt \]
\[ + < d_{k-1}, (H^* - H_{k-1}) d_k > \]
\[ + \frac{1}{2} < d_k, H_k d_k > \]
\[ - f_{k-1}(d_{k-1}, y_{k+1}) \]
\[ - \int_0^1 (1-t) \nabla_x f(x_{k-1} + t(d_{k-1})) \cdot d_{k-1} \, dt \]
\[ - \frac{1}{2} < d_{k-1}, (H^* - H_{k-1}) d_{k-1} > \]  \( (6.37) \)

The above inequality holds for any \( y_k \). Since \( d_{k-1} \) is optimal for \( \Phi_{k-1}(d) \) and, by Lemma 4, \( y_k \) is the maximizer of \( f_{k-1}(d_{k-1}, y) \), we have the inequality
\[ \frac{\partial \Phi_{k-1}(d)}{\partial d} = < \nabla_x f(x_{k-1}, y_k), d_k > + < d_{k-1}, H_{k-1} d_k > \geq 0. \]  \( (6.38) \)
Furthermore, we have

\[ f_{k-1}(d_{k-1}, y_k) = \max_{y \in \mathcal{Y}} f_{k-1}(d_{k-1}, y) \quad (6.39) \]

Inequality (6.37) reduces to

\[
-\frac{1}{2} |d_k|_{H_k}^2 \geq \int_0^1 <d_{k-1}, (H[x_{k-1} + t(d_{k-1})] - H^*)d_1> dt \\
+ <d_{k-1}, (H^* - H_{k-1})d_k > \\
+ \frac{1}{2} <d_k, H_k d_k > . \quad (6.40)
\]

Let \( R_{k-1} \) be defined as

\[
R_{k-1} = -\frac{1}{2} \int_0^1 \{ H[x_{k-1} + t(d_{k-1})] - H^* \} dt + (H^* - H_{k-1}) \} . \quad (6.41)
\]

Inequality (6.40) becomes

\[
-\frac{1}{2} |d_k|_{H_k}^2 \geq -2 <d_{k-1}, (R_{k-1})d_k > + \frac{1}{2} |d_k|_{H_k}^2 \quad (6.42)
\]

or

\[
|d_k|_{H_k}^2 \leq 2 <d_{k-1}, (R_{k-1})d_k > . \quad (6.43)
\]

By (6.6), \( m|d_k|^2 \leq |d_k|_{H_k}^2 \), (6.43) becomes

\[
\frac{|d_k|}{|d_{k-1}|} \leq 2 \frac{|R_{k-1}|}{m} . \quad (6.44)
\]

As \( k \to \infty \), \( |R_{k-1}| \to \frac{1}{2} \frac{|H^* - H_{k-1} (d_{k-1})|}{|d_{k-1}|} \) and (6.44) becomes

\[
\frac{|d_k|}{|d_{k-1}|} \leq 2 \frac{|R_{k-1}|}{m} < r_k . \quad (6.45)
\]
As \( \{x_k\} \to x^* \) and \( \{y_k\} \to y^* \), if (6.24) is satisfied, then \( \{r_k\} \to 0 \). Thus, for some \( K \) we have \( r_k \leq \bar{r} < 1 \), \( \forall k \geq K \) and

\[
|x^* - x_k| = \lim_{t \to \infty} |x_i - x_k| \leq \lim_{t \to \infty} \sum_{j=k}^{t-1} |x_{j+1} - x_j|
\]

\[
\leq r_k |x_k - x_{k-1}| \lim_{t \to \infty} \sum_{j=k}^{t-1} r_j^{j-k}
\]

\[
\leq \frac{r_k}{1 - \bar{r}} |x_k - x_{k-1}|
\]

or

\[
\frac{|x^* - x_k|}{|x_k - x_{k-1}|} \leq \frac{r_k}{1 - \bar{r}} . \tag{6.47}
\]

To prove Q-superlinear convergence, we show that if

\[
\lim_{k \to \infty} \frac{(H^* - H_k)(d_k)}{|d_k|} = 0 \tag{6.48}
\]

then \( \{x_k\} \) is Q-superlinearly convergent.

Inequality (6.47) may be expressed as

\[
|x^* - x_k| \leq \frac{r_k}{1 - \bar{r}} |x_k - x^* + x^* - x_{k-1}|
\]

\[
\frac{|x^* - x_k|}{|x_k - x^*| + |x^* - x_{k-1}|} \leq \frac{r_k}{1 - \bar{r}} . \tag{6.50}
\]

If (6.24) holds, then \( \lim r_k = 0 \) and we have

\[
\lim_{k \to \infty} \frac{|x^* - x_k|}{|x^* - x_{k-1}|} = 0 . \tag{6.51}
\]

This establishes Q-superlinear convergence.

QED.
Appendix

The existence of the Hessian of $\Phi(x)$

The following lemma is a consequence of the results of Wierzbicki[46].

**Lemma 6.** Let Assumptions 1, 2 and 5 hold. Further, let $\lambda_y$ be a real number corresponding to a $y \in Y$ such that $\lambda_y \geq 0$, $\sum_{y \in Y} \lambda_y = 1$.

The Hessian of $\Phi(x)$ denoted by $H_{\Phi(x)}$ exists and is given by

$$H_{\Phi(x)} = \sum_{y \in Y(x)} \lambda_y \nabla^2 f(x, y).$$

**Proof.**

Let the original problem be reformulated as

$$\min_{(x, z) \in \mathbb{R}^{n+1}} z$$

such that

$$f(x, y) - z \leq 0, \forall y \in Y.$$  \hspace{1cm} (A1)

The Lagrangian associated with (A1) is given by

$$L(\lambda, x, z) = z + \sum_{y \in Y} \lambda_y (f(x, y) - z)$$

$$= z(1 - \sum_{y \in Y} \lambda_y) + \sum_{y \in Y} \lambda_y f(x, y)$$

In order to prove the lemma, we need to show that we can find multipliers $\lambda_y$ satisfying Condition A below.

**Condition A:**

$$\begin{cases}
\lambda_y \geq 0, & \sum_{y \in Y(x)} \lambda_y = 1, \ y \in Y(x) \\
\lambda_y = 0, & \ y \in Y \setminus Y(x)
\end{cases}$$

We first consider the first order optimality condition for (A2). At the solution $(\lambda^*, x^*, z^*)$ of (A2), we have
\[ \nabla_\lambda L(\lambda_*, x_*, z_*) = 0 \quad \text{(A4)} \]
\[ \nabla_x L(\lambda_*, x_*, z_*) = 0 \quad \text{(A5)} \]
\[ \nabla_z L(\lambda_*, x_*, z_*) = 0 \quad \text{(A6)} \]

Consider (A4), \( 0 = \nabla_\lambda L(\lambda_*, x_*, z_*) = \sum_{y \in Y} (f(x_*, y) - z_*) \).

For \( y \in Y(x_*) \), we have \( f(x_*, y) - z_* = 0 \).

For \( y \in Y \setminus Y(x_*) \), we have \( f(x_*, y) - z_* < 0 \).

Therefore, \( \nabla_\lambda L(\lambda_*, x_*, z_*) = 0 \) is satisfied for \( \lambda_y \) such that \( y \in Y(x_*) \).

Consider (A5), \( 0 = \nabla_x L(\lambda_*, x_*, z_*) = \sum_{y \in Y} \lambda_y \nabla_x f(x_*, y) \).

For \( x_* \) to be a solution of both (A1) and (A2), and therefore of the original problem, \( x_* \) must also satisfy
\[ 0 \in \partial \Phi(x_*) = \text{conv}\{\nabla_x f(x_*, y) \mid y \in Y(x_*)\} \quad \text{(A7)} \]

From Caratheodory's Theorem (Lemma II.1.1 in Demyanov and Malozemov[11]), there exist \( n+1 \) not necessarily unique \( y \in Y(x_*) \) and corresponding \( \lambda_y \)'s such that
\[ \left\{ \sum \lambda_y \nabla_x f(x_*, y) \mid \lambda_y \geq 0, \sum \lambda_y = 1, y \in Y(x_*) \right\} \quad \text{(A8)} \]

Therefore, using (A8), we can find \( \lambda_y \)'s satisfying (A7), and also satisfying (A5) provided that the \( \lambda_y \)'s meet Condition A.

Consider (A6), \( 0 = \nabla_z L(\lambda_*, x_*, z_*) = 1 - \sum_{y \in Y} \lambda_y \). For \( \lambda_y \)'s that are characterized by (A8), optimality condition (A6) becomes
\[ 0 = \nabla_z L(\lambda_*, x_*, z_*) = 1 - \sum_{y \in Y(x_*)} \lambda_y \quad \text{(A9)} \]

which is also satisfied as \( Y(x_*) \subset Y \).
We have shown that there are \( \lambda_y \)'s satisfying Condition A. We now prove the lemma.

At the solution, the Lagrangian has a Hessian given by

\[
\nabla_{xx}^2 L(\lambda_*, x_*, z_*) = \sum_{y \in Y} \lambda_y \nabla_{xx}^2 f(x_*, y)
\]

(A10)

which is equivalent to

\[
\nabla_{xx}^2 L(\lambda_*, x_*, z_*) = \sum_{y \in Y(x_*)} \lambda_y \nabla_{xx}^2 f(x_*, y)
\]

(A11)

provided that Condition A is satisfied. Without any loss of generality, at any \( x \), Condition A can be satisfied by using Caratheodory's Theorem, and thus

\[
\nabla_{xx}^2 L(\lambda, x, z) = \sum_{y \in Y(x)} \lambda_y \nabla_{xx}^2 f(x, y)
\]

(A12)

Given Condition A, The Lagrangian becomes

\[
L(\lambda, x, z) = \sum_{y \in Y(x)} \lambda_y f(x, y) = z
\]

(A13)

where \( z = \Phi(x) \). Therefore, (A12) becomes

\[
H_{\Phi(x)} = \sum_{y \in Y(x)} \lambda_y \nabla_{xx}^2 f(x, y).
\]

(A14)

QED.

Remarks.
1. The optimal choice of \( \lambda_y \)'s, \( y \in Y(x) \), corresponds to \( \sum_{y \in Y(x)} \lambda_y \nabla_x f(x, y) \in \partial \Phi(x) \) of minimum norm.

2. The relations in Lemma 6 are also used as preconditions for Lemma 2.1 of Kiwiel[19,p276].
3

Numerical Results

1 Introduction

We consider minimax problems where the objective function is the maximum of a smooth function; the max-function is continuous but may not be straightforwardly differentiable. In Chapter 2, we proposed an implementable quasi-Newton algorithm for solving the continuous minimax problem. This consists of a maximization subproblem, a quadratic approximation to the max-function, a quasi-Newton direction and a stepsize strategy of the Armijo type. In this chapter, we discuss the numerical convergence properties of the algorithm and the properties of the quadratic approximation to the max-function. We report numerical results for a number of test problems solved using the proposed algorithm and using another minimax algorithm. We implemented the algorithm developed by Kiwiel [19] as a check to our numerical results. The test problems include published discrete minimax problems.

In Chapter 2, we discussed an algorithm for the problem

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} f(x, y).$$

(1.1)

In this chapter, we briefly present the algorithm along with an alternative algorithm due to Kiwiel. We then discuss numerical experience using the algorithms for test problems

\footnote{A discussion of this stepsize strategy can be found in Luenberger [22].}
of specific categories. Kiwiel's algorithm is used more to substantiate the results than necessarily as a comparison. When the function $f(x,y)$ is convex in both $x$ and $y$, with $y$ bounded above and below, the continuous minimax problem can be expressed as a discrete minimax problem. We have used the algorithms in such cases and compared the results with those of a nonlinear programming formulation of the discrete minimax using the NAG library optimization subroutine E04VDF.

This chapter is structured as follows. In Section 1.1, we give the definitions of the main functions used in the algorithm. In Section 1.2, we present the proposed quasi-Newton algorithm and in Section 1.3, Kiwiel's algorithm. In Section 2, we present the numerical results for 20 test problems.

1.1 Definitions

In this section, we give the main concepts used in the proposed quasi-Newton algorithm given in Section 1.2. The same concepts are referred to in Section 1.3 where Kiwiel's algorithm is given.

To describe the algorithms for solving (1.1), we first introduce some notation. Let

$$\Phi(x) = \max_{y \in Y} f(x,y)$$  \hspace{1cm} (1.2)

which we call the max-function.

$$\Phi_k(d) = \max_{y \in Y} f_k(d,y)$$ \hspace{1cm} (1.3)

where $f_k(d,y)$, the quadratic approximation to $f(x,y)$, is

$$f_k(d,y) = f(x_k,y) + \langle \nabla_x f(x_k,y), d \rangle + \frac{1}{2}d^2_{H_k}$$ \hspace{1cm} (1.4)

---

2The definitions (Section 1.1) and the algorithm (Section 1.2) have been presented in Chapter 2. Kiwiel's algorithm has been presented in Appendix 3 of Chapter 1. These are re-introduced here for ease of reference.
where \(|d|_{H_k}^2 = \langle d, H_k d \rangle\) and \(H_k\) is a positive definite Hessian approximation to the Hessian with respect to \(x\) of \(\Phi(x)\) at the \(k^{th}\) iteration. \(H_k\) is used in the proposed quasi-Newton algorithm. Kiwiel used a linear approximation: the term \(|d|_{H_k}^2\) is not used.

Let
\[
Y(x) = \left\{ y \in Y \mid y = \arg \max_{y \in Y} f(x, y) \right\}.
\]

We write the initial problem as
\[
\min_{x \in \mathbb{R}^n} \Phi(x). \tag{1.6}
\]

To solve Problem (1.6), we construct the iterative sequence \(x_{k+1} = x_k + \alpha_k d_k\), where \(\alpha_k\) is calculated according to a certain rule based on the Armijo stepsize strategy while \(d_k\) is the solution of the auxiliary problem
\[
\min_{d \in \mathbb{R}^n} \Phi_k(d). \tag{1.7}
\]

Corresponding to the solution \(d_k\) of Problem (1.7), we have the set of maximizers given by
\[
Y_{k+1} = \left\{ y_{k+1} \in Y \mid y_{k+1} = \arg \max_{y \in Y} f_k(d_k, y) \right\}. \tag{1.8}
\]

Finally, we define the approximate directional derivative as
\[
\Psi_k = \min_{d \in \mathbb{R}^n} \Phi_k(d) - \Phi(x_k). \tag{1.9}
\]
Assumptions
1. \( Y \subset R^n \) is convex and compact.
2. \( f(x,y) \) is twice continuously differentiable both in \( x \) and \( y \).
3. There exists a number \( m > 0 \) such that \( m|x|^2 \leq |x|^2_{H_1} \) for all \( x \in R^n \).
4. In the neighborhood of the solution \( x* \) of (1.6), there exists \( b > 0 \) such that \( \forall x \in \{ x | |x - x*| < b, x \in R^n \} \) the Hessian with respect to \( x \) of \( f(x,y) \) for all \( y \in Y \) is positive definite.\(^3\)
5. \( \Phi(x) \) is bounded from below in \( R^n \).

1.2 The Proposed Quasi-Newton Algorithm

In this section, we give the proposed quasi-Newton algorithm for solving Problem (1.6). The algorithm includes a search direction finding subproblem where we maximize the quadratic approximation to the max-function that includes a linear penalty term.\(^4\) It has been shown in Chapter 2 that the use of a squared penalty term does not affect the properties of the algorithm. We tested the algorithm using both the squared and the non-squared penalty terms.

The Algorithm
Step 0. Initialization.
Select \( x_1, y_1, H_1 \).
Select final accuracy \( \epsilon \geq 0 \).
Select linesearch parameter \( c, 0 < c < 1 \).
Select stepsizes parameter \( \lambda, 0 < \lambda < 1 \).
Select penalty function coefficient \( C, 0 < C < \infty \).
Set \( k=1 \).

\(^3\)This is required for demonstrating convergence to unit stepsize and local superlinear convergence.
\(^4\)Shanno[44] discuss practical issues on the use of penalty functions.
Step 1. Maximization at current point \( x_k \).
Solve \( \Phi(x_k) = \max_{y \in Y} f(x_k, y) \). \( (1.10) \)

Step 2. Direction-finding subproblem.
Find \( y_{k+1} \) from
\[
y_{k+1} = \arg \max_{y \in Y} \left\{ f(x_k, y) - \frac{1}{2} \left| \nabla_x f(x_k, y) \right|^2_{H_k} - C(\Phi(x_k) - f(x_k, y)) \right\}.
\] \( (1.11) \)
Set:
\[
d_k = -H_k^{-1} \nabla_x f(x_k, y_{k+1}) \quad \Psi_k = f(x_k, y_{k+1}) - \frac{1}{2} \left| \nabla_x f(x_k, y_{k+1}) \right|^2_{H_k} - \Phi(x_k)
\] \( (1.12) \quad (1.13) \)

If \( \Psi_k \geq -\epsilon \), then terminate. \( (1.14) \)

Step 3. Line search.
Find \( \alpha_k \) from
\[
\alpha_k = \max \{ \alpha \mid \Phi(x_k + \alpha d) - \Phi(x_k) \leq c \alpha \Psi_k, \alpha = (\lambda)^i, i = 0, 1, 2, \ldots \}
\] \( (1.15) \)
Set:
\[
x_{k+1} = x_k + \alpha_k d_k
\]
\[
k = k + 1
\]
Update the Hessian\(^5\).
Go to Step 1.

1.3 Kiwiel's Algorithm

In this section we give the algorithm developed by Kiwiel\([19]\) for solving \( (1.6) \). The algorithm includes a search direction finding subproblem where a linear approximation to the \( f(x,y) \) is maximized. The resulting maximizer is used to find a subgradient at \( x_k \).

Kiwiel's algorithm uses a convex combination of subgradients to define a direction of descent. Global convergence and the rate of convergence are discussed in Kiwiel\([19]\).

\(^5\)In the implementation of the proposed algorithm, we used the BFGS updating formula (see Luenberger\([22]\)). A discussion on the local convergence of quasi-Newton methods can be found in Boggs, Tolle and Wang\([2]\), and Rustem\([39]\).
The Algorithm

Step 0. Initialization.

Select $x_1, y_i$.
Select final accuracy $\varepsilon \geq 0$.
Select linesearch parameter $c$, $0 < c < 1$.
Select stepsize parameter $\lambda$, $0 < \lambda < 1$.
Select a linear approximation parameter $m$, $0 < m < 1$.
Set $k = 1$.

Step 1. Maximization at current point $x_k$.
Solve $\Phi(x_k) = \max_{y \in Y} \{f(x_k, y)\}$. (1.16)

Step 2. Direction-finding subproblem.
Set $x = x_k$ and use the direction-finding algorithm
with the parameters $\varepsilon \geq 0$ and $m$
until it stops, returning $d_k$ and $\Psi_k$.
If $\Psi_k \geq -\varepsilon$, then terminate.

Step 3. Line search.
Find $\alpha_k$ from
$\alpha_k = \max\{\alpha \mid \Phi(x_k + \alpha d) - \Phi(x_k) \leq c\alpha \Psi_k, \alpha = (\lambda)^i, i = 0, 1, 2, \ldots\}$ (1.17)
Set:
$x_{k+1} = x_k + \alpha_k d_k$
$k = k + 1$
Go to Step 1.
Direction-Finding Algorithm (Step 2 of Kiwiel's algorithm)

Data: \( x_k \in \mathbb{R}^n \), \( \Phi(x_k) \), \( \epsilon_x \geq 0 \), \( m \in (0,1) \).

Step 0. Initialization.

Set \( x = x_k \) and \( \Phi(x_k) = \Phi(x) \).

Select any \( y \in Y \) and set \( p_0 = \nabla_x f(x, y_0), \Theta_0 = f(x, y_1) \).

Set \( t = 1 \).

Step 1. Reduced dual solving.

Find a number \( \mu_t \) to solve

\[
\min_{\mu \in \mathbb{R}} \left\{ \frac{1}{2} \left(1 - \mu\right) p_{t-1} + \mu \nabla_x f(x, y_t) \right\}^2 - (1 - \mu) \Theta_{t-1} - \mu f(x, y_t) \}.
\]

Set

\[
p_t = (1 - \mu_t) p_{t-1} + (\mu_t) \nabla_x f(x, y_t)
\]

\[
\Theta_t = (1 - \mu_t) \Theta_{t-1} + (\mu_t) f(x, y_t)
\]

\[
\Psi_t = -\left\{ p_t^2 + \Phi(x) - \Theta_t \right\}
\]

If \( \Psi_t \geq -\epsilon_x \) then terminate returning \( d_k = -p_t \) and \( \Psi_k = \Psi_t \).

(1.23)

Otherwise, proceed.

Step 2. Primal optimality testing.

Set \( d_t = -p_t \).

Find

\[
y_{t+1} = \arg \max_{y \in Y} \left\{ f(x, y) + \langle \nabla_x f(x, y), d_t \rangle \right\}.
\]

If \( f(x, y_{t+1}) + \langle \nabla_x f(x, y_{t+1}), d_t \rangle > -\Phi(x) \leq m \Psi_t \)

then terminate returning \( d_k = -p_t \) and \( \Psi_k = \Psi_t \).

Otherwise set \( t = t + 1 \) and go to step 1.
2 Numerical Results

Test problems have been grouped according to the convexity in the $x$-space and concavity in the $y$-space of the objective function. We present results for different groups in the following sections:

- **Section 2.2:** 10 convex-concave problems
- **Section 2.3:** 5 convex-convex problems
- **Section 2.4:** 5 published discrete minimax problems

2.1 Implementation, Output Variables and the Stopping Criterion

The proposed quasi-Newton algorithm and Kiwiel's[19] algorithm both require a maximization subproblem to be solved in Step 1. Because this is an optimization problem within the main body of the algorithms, both algorithms are relatively more computationally expensive compared to most nonlinear programming algorithms. For the maximization subproblem we used the comprehensive NAG optimization routine E04VDF which can handle both linear and nonlinear constraints of a nonlinear programming problem.

We implemented Kiwiel's algorithm as a check to the solutions found by the proposed algorithm. All the test problems given in this paper were solved by using both algorithms. Some of the test problems can be reformulated as discrete minimax problems (Test Problems 11-15) and some are originally discrete minimax problems (Test Problems 16-20). The test problems which have discrete minimax formulations were further reformulated as a nonlinear programming problem and solved using NAG E04VDF as a further check to the solutions found by the two minimax algorithms.
In the implementation of Kiwiel's algorithm we set the linear approximation parameter $m$ to a constant for all the test problems: $m = 2.0\times10^{-4}$.

Output Variables:

- **Epsilon $\varepsilon$**: termination parameter where $0 < \varepsilon << 1$
- **Lambda $\lambda$**: stepsize parameter where $0 < \lambda < 1$
- **Line coefficient c**: line search parameter
- **Penalty coefficient $C$**: parameter in the direction-finding subproblem where $0 << C < \infty$
- **$x_0$**: initial value of $x$
- **$y_0$**: initial value of $y$
- **$\|d\|$**: norm of the direction $d$ at termination
- **$\Psi$**: approximate directional derivative at termination
- **$\Phi(x_*)$**: objective function value at the solution
- **$x_*$**: final value of $x$
- **$y_*$**: final value of $y$
- **$k_\alpha l\alpha_\alpha = 1; \forall k \geq k_\alpha$**: iteration number where the stepsize equals 1 for all succeeding iterations
- **No. of iterations**: total number of iterations the algorithm took to solve the problem
- **Time**: total computer time using a Sun Sparcstation ELC running at 7 million instructions per second

Some of the output variables given above are either not part of the standard output of NAG E04VDF or not applicable to the nonlinear programming problem. For these variables the notation **na** is used to indicate either non-availability of output or non-applicability to the problem formulation.
The Stopping Criterion

The stopping criterion for both the proposed algorithm and Kiwiel's [19] algorithm is the condition that the approximate directional derivative is sufficiently close to zero, i.e.

\[ \text{If } \Psi_k \geq -\varepsilon, \text{ then terminate.} \] (2.1)

The values of the stopping parameter \( \varepsilon \) used in the test problems are within the range \([1.0e^{-6}, 1.0e^{-14}]\). In cases where the Hessian approximation \( H_k \) is fairly accurate, this condition is very similar to a stopping criterion based on the norm of the search direction \( d_k \). However, (2.1) is more appropriate when the Hessian approximation is not sufficiently accurate.

2.2 Convex-Concave Problems

Test Problems 1-9 are all convex in the \( x \)-space and concave in the \( y \)-space. These problems illustrate the performance of the proposed quasi-Newton algorithm when the max-function has a unique \( y \) maximizer for a fixed \( x \). We also solved the problems using Kiwiel's [19] algorithm as a check to the solutions found by the proposed algorithm.

Test Problem 1 illustrates the performance of the proposed algorithm for three values of the penalty coefficient, \( 1.0e^6 \), \( 1.0e^4 \) and \( 1.0e^3 \). The use of the penalty coefficient of \( 1.0e^6 \) gave the best performance in terms of time and number of iterations to get to the optimal solution. For the rest of the test problems, Problems 2-20, the penalty coefficient has been fixed at \( 1.0e^6 \).

Test Problem 10 is nonconvex in the \( x \)-space and nonconcave in the \( y \)-space. Both algorithms found a minimax solution where \( \text{Hessian}_{xx} \) is positive definite and \( \text{Hessian}_{yy} \) is negative definite. We have included it in this section as it is convex-concave at the solution.
Problem 1

\( x \in \mathbb{R}^2, \ y \in \mathbb{R}^2 \)

Objective function:

\[
\begin{align*}
f &= 5x_0^2 + 5x_1^2 - y_0^2 - y_1^2 \\
&\quad + x_0(-y_0 + y_1 + 5) \\
&\quad + x_1(y_0 - y_1 + 3)
\end{align*}
\]

\(-5 \leq y_i \leq 5, \ i = 0, 1\)

Penalty coefficient C: 1.000000 e+06

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-11</td>
<td>1.000000 e-11</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient c</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient C</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
</tbody>
</table>

\( x_0 \)

<table>
<thead>
<tr>
<th></th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>-10.0</td>
<td>-10.0</td>
<td></td>
<td>-10.0</td>
</tr>
</tbody>
</table>

\( y_0 \)

<table>
<thead>
<tr>
<th></th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>5.0</td>
<td></td>
<td>5.0</td>
</tr>
<tr>
<td>-5.0</td>
<td>-5.0</td>
<td></td>
<td>-5.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Idl</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.453185 e-07</td>
<td>4.132408 e-07</td>
<td></td>
<td>2.322320 e-05</td>
</tr>
<tr>
<td>-1.423084 e-12</td>
<td>-9.330314 e-13</td>
<td></td>
<td>-9.519834 e-10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \Phi(x*) )</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.683333 e+00</td>
<td>-1.683333 e+00</td>
<td></td>
<td>-1.683333 e+00</td>
</tr>
<tr>
<td>-4.833337 e-01</td>
<td>-4.833335 e-01</td>
<td></td>
<td>-4.833313 e-01</td>
</tr>
<tr>
<td>-3.166664 e-01</td>
<td>-3.166671 e-01</td>
<td></td>
<td>-3.166682 e-01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( y* )</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8.333333 e-02</td>
<td>8.333322 e-02</td>
<td></td>
<td>8.333152 e-02</td>
</tr>
<tr>
<td>-8.333333 e-02</td>
<td>-8.333322 e-02</td>
<td></td>
<td>-8.333152 e-02</td>
</tr>
</tbody>
</table>

\( k_a l \alpha \_k = 1; \forall k \geq k_a \)

<table>
<thead>
<tr>
<th></th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of iterations</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>11</td>
<td></td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 sec</td>
<td>0.5 sec</td>
<td></td>
<td>1.1 sec</td>
</tr>
</tbody>
</table>
Penalty coefficient C: 1.000000 e+04

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwieli's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>Epsilon ε</td>
<td>1.000000 e-11</td>
<td>1.000000 e-11</td>
</tr>
<tr>
<td>Lambda λ</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient c</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient C</td>
<td>1.000000 e+04</td>
<td>1.000000 e+04</td>
</tr>
</tbody>
</table>

| x₀          | 10.0 | 10.0 | 10.0 |
|            | -10.0 | -10.0 | -10.0 |
| y₀          | 5.0  | 5.0  | 5.0  |
|            | -5.0  | -5.0  | -5.0  |
| ld| 5.979549 e-07 | 5.341083 e-07 | 2.322320 e-05 |
| Ψ           | -9.684253 e-12 | -1.787015 e-12 | -9.519834 e-10 |
| Φ(xₙ)       | -1.683333 e+00 | -1.683333 e+00 | -1.683333 e+00 |
| xₙ          | -4.833322 e-01 | -4.833338 e-01 | -4.833313 e-01 |
|            | -3.166692 e-01 | -3.166663 e-01 | -3.166682 e-01 |
| yₙ          | 8.333192 e-02  | 8.333374 e-02  | 8.333152 e-02  |
|            | -8.333192 e-02 | -8.333374 e-02 | -8.333152 e-02 |

\[ k_a \mid x_k = 1; \forall k \geq k_a \mid 4 \quad 18 \quad 0 \]

| No. of iterations | 20 | 19 | 23 |
| Time              | 0.7 sec | 0.9 sec | 1.1 sec |
Penalty coefficient C: 1.000000 e+03

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>Epsilon $\epsilon$</td>
<td>1.000000 e-11</td>
<td>1.000000 e-11</td>
</tr>
<tr>
<td>Lambda $\lambda$</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient $c$</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient C</td>
<td>1.000000 e+03</td>
<td>1.000000 e+03</td>
</tr>
</tbody>
</table>

| $x_0$                      | 10.0               | 10.0              | 10.0               |
|                            | -10.0              | -10.0             | -10.0              |
| $y_0$                      | 5.0                | 5.0               | 5.0                |
|                            | -5.0               | -5.0              | -5.0               |

| $|d|$ | 8.770758 e-07 | 6.618906 e-07 | 2.322320 e-05 |
| $\Psi$ | -6.965761 e-12 | -2.942757 e-12 | -9.519834 e-10 |
| $\Phi(x)$ | -1.683333 e+00 | -1.683333 e+00 | -1.683333 e+00 |

| $x$ | -4.833345 e-01 | -4.833327 e-01 | -4.833313 e-01 |
|     | -3.166678 e-01 | -3.166670 e-01 | -3.166682 e-01 |
| $y$ | 8.333322 e-02  | 8.333283 e-02  | 8.333152 e-02  |
|     | -8.333322 e-02 | -8.333283 e-02 | -8.333152 e-02 |

$k_a | \alpha_k = 1; \forall k \geq k_a$ | 0 | 34 | 0 |
No. of iterations | 24 | 36 | 23 |
Time | 1.0 sec | 1.4 sec | 1.1 sec |
Problem 2

\( x \in \mathbb{R}^1, \ y \in \mathbb{R}^1 \)

Objective function:

\[
 f = 2(x_0 - 2)x_0 + x_0^2y_0 - 3(y_0 - 1)y_0
\]

\(-3 \leq y_0 \leq 3\)

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-11</td>
<td>1.000000 e-11</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
<tr>
<td>( x_0 )</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>( y_0 )</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>( l_1 )</td>
<td>1.733550 e-06</td>
<td>2.144992 e-08</td>
</tr>
<tr>
<td>( \Psi )</td>
<td>-8.392731 e-12</td>
<td>-1.332268 e-15</td>
</tr>
<tr>
<td>( \Phi(x) )</td>
<td>-8.184562 e-01</td>
<td>-8.184562 e-01</td>
</tr>
<tr>
<td>( x^* )</td>
<td>7.696089 e-01</td>
<td>7.696106 e-01</td>
</tr>
<tr>
<td>( y^* )</td>
<td>5.987167 e-01</td>
<td>5.987167 e-01</td>
</tr>
<tr>
<td>( k_a</td>
<td>\alpha_k = 1; \forall k \geq k_a )</td>
<td>2</td>
</tr>
<tr>
<td>No. of iterations</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Time</td>
<td>0.2 sec</td>
<td>0.2 sec</td>
</tr>
</tbody>
</table>

78
**Problem 3**

\[ x \in \mathbb{R}^1, \ y \in \mathbb{R}^1 \]

**Objective function:**

\[ f = -4x_0^2 - y_0^2 + x_0^2y_0 \]

\[ -5 \leq y \leq 5 \]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-07</td>
<td>1.000000 e-07</td>
<td>1.000000 e-07</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.2 sec</td>
<td>0.3 sec</td>
</tr>
</tbody>
</table>

|        | 5.0  | 5.0  | 5.0  |
|        |      |      |      |
|        | 5.0  | 5.0  | 5.0  |
|        |      |      |      |
|        | 3.171570 e-07 | 2.202170 e-10 | 5.101749 e-07 |
|        | -8.224532 e-13 | 0.000000 e+00 | -9.380940 e-12 |
|        | -1.600000 e+01 | -1.600000 e+01 | -1.600000 e+01 |
|        | 2.828427 e+00 | 2.828427 e+00 | 2.828427 e+00 |
|        | 4.000000 e+00 | 4.000000 e+00 | 4.000001 e+00 |

|        | 3    | 3    | 1    |
|        |      |      |      |
|        |      |      |      |
|        |      |      |      |
Problem 4

\( x \in \mathbb{R}^2, y \in \mathbb{R}^2 \)

Objective function:

\[
f = 4x_0^2 + y_0x_0^2 + 4x_1^2 + y_1x_1^2 - 2y_0^2 - y_1^2
+ x_0(5 - y_0 + y_1) + x_1(3 + y_0 - y_1)
\]

\(-5 \leq y_i \leq 5, \ i = 0, 1\)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Proposed Algorithm</th>
<th>Kiwieli's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-11</td>
<td>1.000000 e-11</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient c</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient C</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
</tbody>
</table>

| \( x_0 \) | 4.0 | 4.0 | 4.0 |
| \( y_0 \) | 5.0 | 5.0 | 5.0 |

\( l d l \)

\[
-9.091615 \times 10^{-8} \quad -8.520130 \times 10^{-8} \quad 3.774084 \times 10^{-6}
\]

\( \Psi \)

\[
-8.258727 \times 10^{-12} \quad -6.776801 \times 10^{-12} \quad -8.857803 \times 10^{-12}
\]

\( \Phi(x_*) \)

\[
-2.081315 \times 10^{00} \quad -2.081315 \times 10^{00} \quad -2.081315 \times 10^{00}
\]

\( x_* \)

\[
-5.865187 \times 10^{-01} \quad -5.865187 \times 10^{-01} \quad -5.865187 \times 10^{-01}
-3.955779 \times 10^{-01} \quad -3.955776 \times 10^{-01} \quad -3.955753 \times 10^{-01}
\]

\( y_* \)

\[
1.337365 \times 10^{-01} \quad 1.337363 \times 10^{-01} \quad 1.337391 \times 10^{-01}
-1.723010 \times 10^{-02} \quad -1.722975 \times 10^{-02} \quad -1.723431 \times 10^{-02}
\]

\( k_a \mid \alpha_k = 1; \forall k \geq k_a \)

\[
3 \quad 3 \quad 0
\]

No. of iterations

\[
24 \quad 25 \quad 20
\]

Time

\[
0.8 \text{ sec} \quad 1.4 \text{ sec} \quad 1.0 \text{ sec}
\]
Problem 5

\[ x \in \mathbb{R}^2, \ y \in \mathbb{R}^3 \]

Objective function:

\[
f = y_0^2(x_1^2 + 10x_0 - 10) + 5y_2x_1 + y_1^2(x_0^2 - 6x_1 - 10) + \frac{1}{10} y_0x_0 + y_2^2(x_1(x_0 + 8) - 10) + \frac{1}{10} y_1x_0
\]

\[-3 \leq y_i \leq 3, \ i = 0, 1, 2\]

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwiels Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ε (\varepsilon)</td>
<td>1.000000 e-11</td>
<td>1.000000 e-11</td>
</tr>
<tr>
<td>Lambda λ (\lambda)</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient (c)</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient (C)</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwiels Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_0)</td>
<td>-1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>(y_0)</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(y_0)</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(y_0)</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

| \(ldl\)                         | 1.900991 e-05      | 1.640312 e-05     | 3.154115 e-06    |
| \(\Psi\)                        | -7.283122 e-12     | -7.086279 e-12    | -9.264988 e-12   |
| \(\Phi(x_\star)\)               | 2.243242 e-02      | 2.243247 e-02     | 2.243241 e-02    |

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwiels Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_\star)</td>
<td>-1.590179 e-01</td>
<td>-1.588411 e-01</td>
</tr>
<tr>
<td>(y_\star)</td>
<td>1.000813 e+00</td>
<td>1.000700 e+00</td>
</tr>
<tr>
<td>(y_\star)</td>
<td>-2.436165 e-02</td>
<td>-2.436401 e-02</td>
</tr>
<tr>
<td>(y_\star)</td>
<td>-3.178740 e-02</td>
<td>-3.178816 e-02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters: (k_a) (\alpha_k = 1; \forall k \geq k_a)</th>
<th>Proposed Algorithm</th>
<th>Kiwiels Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>No. of iterations</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td>Time</td>
<td>0.8 sec</td>
<td>1.9 sec</td>
</tr>
</tbody>
</table>

81
Problem 6

\( x \in \mathbb{R}^3, \ y \in \mathbb{R}^3 \)

Objective function:

\[
f = 2x_0^2 + 3x_1^2 + x_2^2 - y_0^2 - y_1^2 - y_2^2 \\
+ (1 - x_0^2) y_0 + (x_1^2 - 1) y_1 + (x_2^2 - 1) y_2
\]

\(-1 \leq y_i \leq 1, \ i = 0, 1, 2\)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Proposed Algorithm</th>
<th>Kiwiell's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-11</td>
<td>1.000000 e-11</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( x_0 )</th>
<th>-2.0</th>
<th>-2.0</th>
<th>-2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>-2.0</td>
<td>-2.0</td>
<td>-2.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( y_0 )</th>
<th>1.0</th>
<th>1.0</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

| \( \lambda_1 \) | 7.939981 e-07 | 9.202250 e-07 | 1.303676 e-06 |
| \( \Psi \)      | -8.761880 e-12 | -8.174794 e-12 | -9.421797 e-12 |
| \( \Phi(x^*) \) | -6.358778 e+00 | -6.358778 e+00 | -6.358778 e+00 |

| \( x^* \) | 1.542470 e+00 | 1.542471 e+00 | 1.542469 e+00 |
|           | -2.315483 e+00 | -2.315482 e+00 | -2.315479 e+00 |
|           | -2.897541 e+00 | -2.897542 e+00 | -2.897537 e+00 |

| \( y^* \) | 1.896066 e-01 | 1.896087 e-01 | 1.896056 e-01 |
|           | 7.319584 e-01 | 7.319577 e-01 | 7.319556 e-01 |
|           | 1.263096 e-01 | 1.263098 e-01 | 1.263017 e-01 |

| \( k_\alpha \) | 6 | 6 | 28 |
| No. of iterations | 74 | 71 | 31 |
| Time           | 2.9 sec | 4.5 sec | 2.9 sec |
Problem 7

\( x \in \mathbb{R}^4, y \in \mathbb{R}^3 \)

Objective function:

\[
\begin{align*}
  f &= -y_0^2 + y_0(x_0^2 - x_1 + x_2 - x_3 + 2) \\
  &\quad - y_1^2 + y_1(-x_0 + 2x_1^2 - x_2^2 + 2x_3 + 1) \\
  &\quad - y_2^2 + y_2(2x_0 - x_1 + 2x_2 - x_3^2 + 5) \\
  &\quad + 5x_0^2 + 4x_1^2 + 3x_2^2 + 2x_3^2 \\
  -2 \leq y_i \leq 2, \quad i = 0, 1, 2
\end{align*}
\]

Parameters:

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Epsilon ( \varepsilon )</th>
<th>Lambda ( \lambda )</th>
<th>Line coefficient ( c )</th>
<th>Penalty coefficient ( C )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Algorithm</td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
<td>Kiwiell's Algorithm</td>
<td></td>
</tr>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
<td></td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td></td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td></td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td></td>
</tr>
</tbody>
</table>

\( x_0 \):

<table>
<thead>
<tr>
<th>-2.0</th>
<th>2.0</th>
<th>-2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>-2.0</td>
<td>-2.0</td>
<td>-2.0</td>
</tr>
<tr>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

\( y_0 \):

<table>
<thead>
<tr>
<th>2.0</th>
<th>2.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

\( 1dF \):

| 2.503993 e-05 | 1.087402 e-04 | 1.972328 e-04 |

\( \Psi \):

| -6.639361 e-09 | -6.127507 e-09 | -5.146115 e-09 |

\( \Phi(x*) \):

| 4.542970 e+00 | 4.542970 e+00 | 4.542970 e+00 |

\( x* \):

<table>
<thead>
<tr>
<th>-2.315354 e-01</th>
<th>-2.315446 e-01</th>
<th>-2.315577 e-01</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.227950 e-01</td>
<td>2.227998 e-01</td>
<td>2.228115 e-01</td>
</tr>
<tr>
<td>-6.755253 e-01</td>
<td>-6.755340 e-01</td>
<td>-6.755257 e-01</td>
</tr>
<tr>
<td>-8.376421 e-02</td>
<td>-8.374315 e-02</td>
<td>-8.375179 e-02</td>
</tr>
</tbody>
</table>

\( y* \):

<table>
<thead>
<tr>
<th>6.195319 e-01</th>
<th>6.195111 e-01</th>
<th>6.195168 e-01</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.535066 e-01</td>
<td>3.534958 e-01</td>
<td>3.535046 e-01</td>
</tr>
<tr>
<td>1.477986 e+00</td>
<td>1.478015 e+00</td>
<td>1.478004 e+00</td>
</tr>
</tbody>
</table>

\( k_\alpha | \alpha_k = 1; \forall k > k_\alpha \)

| 27 | 17 | 0 |

No. of iterations

| 28 | 18 | 22 |

Time

| 1.5 sec | 2.1 sec | 1.6 sec |
Problem 8

\[ x \in \mathbb{R}^5, \ y \in \mathbb{R}^2 \]

Objective function:

\[
f = 2x_4x_0 + 3x_3x_1 + x_4x_2 + 5x_3^2 + 5x_2^2 - y_0^2 - y_1^2 - x_3(y_3 - y_4 - 5)
+ x_4(y_3 - y_4 + 3) - \frac{y_0^2}{2} - \frac{y_1^2}{2} + \frac{(x_0^2 - 1)y_0 + (x_1^2 - 1)y_1 + (x_2^2 - 1)y_2 - y_4^2}{2}
- 3 \leq y_i \leq 3, \ i = 0, 1, 2, 3, 4
\]

Proposed Algorithm

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th>Kiwiell's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000E-08</td>
<td>1.000E-08</td>
<td>1.000E-08</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000E-01</td>
<td>5.000E-01</td>
<td>5.000E-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000E-04</td>
<td>1.000E-04</td>
<td>1.000E-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000E+00</td>
<td>1.000E+00</td>
<td>1.000E+00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposed Algorithm</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_0 )</td>
<td>-4.0</td>
<td>-4.0</td>
</tr>
<tr>
<td>( y_0 )</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kiwiell's Algorithm</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x^* )</td>
<td>1.425208E+00</td>
<td>1.425180E+00</td>
</tr>
<tr>
<td>( y^* )</td>
<td>5.156131E-01</td>
<td>5.155692E-01</td>
</tr>
<tr>
<td>( k_\alpha \mid k_\alpha ) = 1; \ \forall k &gt; k_\alpha )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No. of iterations</td>
<td>72</td>
<td>43</td>
</tr>
<tr>
<td>Time</td>
<td>8.8 sec</td>
<td>8.4 sec</td>
</tr>
</tbody>
</table>
Problem 9  
\( x \in \mathbb{R}^2, \ y \in \mathbb{R}^2 \)

Objective function:
\[
f = 4(x_0 - 2)^2 - 2y_0^2 + x_0^2y_0 - y_i^2 + 2x_i^2y_i
\]

\(-5 \leq y_i \leq 5, \ i = 0,1\)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Proposed Algorithm</th>
<th>Kiwiell's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
<tr>
<td>( x_0 )</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>( y_0 )</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>1d1</td>
<td>7.327505 e-05</td>
<td>1.383270 e-04</td>
</tr>
<tr>
<td>( \Psi )</td>
<td>-2.776592 e-09</td>
<td>-9.567186 e-09</td>
</tr>
<tr>
<td>( \Phi(x*) )</td>
<td>1.403883 e+00</td>
<td>1.403884 e+00</td>
</tr>
<tr>
<td>( x* )</td>
<td>1.695411 e+00</td>
<td>1.695410 e+00</td>
</tr>
<tr>
<td></td>
<td>9.742972 e-01</td>
<td>9.684607 e-01</td>
</tr>
<tr>
<td>( y* )</td>
<td>7.186165 e-01</td>
<td>7.186042 e-01</td>
</tr>
<tr>
<td>( k_{\alpha_{k}} = 1; \forall k &gt; k_{\alpha} )</td>
<td>0</td>
<td>194</td>
</tr>
<tr>
<td>No. of iterations</td>
<td>34</td>
<td>195</td>
</tr>
<tr>
<td>Time</td>
<td>1.4 sec</td>
<td>12.2 sec</td>
</tr>
</tbody>
</table>
Problem 10

\( x \in \mathbb{R}^2, y \in \mathbb{R}^2 \)

Objective function:

\[
f = x_0^4y_1 + 2x_0^3y_0 - x_1^2y_1(y_1 - 3) - 2x_1(y_0 - 3)^2
\]

\[-0 \leq y_i \leq 3, \quad i = 0, 1\]

<table>
<thead>
<tr>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Squared Penalty Term</strong></td>
<td><strong>Non-Squared Penalty Term</strong></td>
</tr>
<tr>
<td>Epsilon ε</td>
<td>1.000000 e-11</td>
</tr>
<tr>
<td>Lambda λ</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient c</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient C</td>
<td>1.000000 e+06</td>
</tr>
<tr>
<td>( x_0 )</td>
<td>10.0</td>
</tr>
<tr>
<td>( y_0 )</td>
<td>3.0</td>
</tr>
<tr>
<td>( l_d l )</td>
<td>2.820707 e-07</td>
</tr>
<tr>
<td>( \Psi )</td>
<td>-3.824940 e-12</td>
</tr>
<tr>
<td>( \Phi(x_0) )</td>
<td>-2.468775 e+00</td>
</tr>
<tr>
<td>( x_1 )</td>
<td>-1.180674 e+00</td>
</tr>
<tr>
<td>( y_1 )</td>
<td>9.128248 e-01</td>
</tr>
<tr>
<td>( k_\alpha</td>
<td>\alpha_k = 1; \forall k \geq k_\alpha )</td>
</tr>
<tr>
<td>No. of iterations</td>
<td>33</td>
</tr>
<tr>
<td>Time</td>
<td>0.9 sec</td>
</tr>
</tbody>
</table>
2.3 Convex-Convex Problems

Test Problems 11-15 are all convex in both the $x$-space and the $y$-space. These problems illustrate the performance of the proposed quasi-Newton algorithm when the max-function may have multiple maximizers for a fixed $x$. The set of maximizers comprises the extreme points in the $y$-space defined by the upper and lower bounds on $y$. We also solved the problems using Kiwiel's[19] algorithm as well as the NAG optimization subroutine E04VDF as a check to the solutions found by the proposed algorithm.

Convex-convex continuous minimax problems subject to bounds constraints in the $y$-space can be reformulated as a discrete minimax problem using the following transformation.

**Continuous minimax formulation:**

\[
\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} f(x, y) \\
y_i \leq y \leq y_u, \quad i = 1, \ldots, m
\]  

(2.2)

**Discrete minimax formulation:**

\[
\min_{x \in \mathbb{R}^n} \max_{t \in T} \{f_t(x, y_t)\}
\]

(2.3)

where

\[
t \in T = \{1, \ldots, 2^n\}
\]

The set $T$ is the index set of extreme points and $y_t$ is the $y$-variable instantiated to the extreme point $t$, $t \in T$. 

87
The discrete minimax formulation is equivalent to the nonlinear programming problem in
\( n+1 \) variables \( x \in \mathbb{R}^n \) and \( z \in \mathbb{R}^1 \).

**Nonlinear programming formulation:**

\[
\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^1} z \quad \text{subject to} \quad f_t(x, y_t) - z \leq 0, \quad t = 1, ..., 2^m.
\]

At the solution of Problem (2.4) the values of \( f_t \) correspond to the values of the
continuous minimax formulation at the extreme points of \( y \). Problem (2.4) has been
solved using the NAG nonlinear programming subroutine E04VDF.

For Test Problems 11-15, we report the value of the objective function at the extreme
points found using our proposed algorithm, Kiwiels algorithm and NAG E04VDF.
Problem 11

\( x \in \mathbb{R}^1, \ y \in \mathbb{R}^1 \)

Objective function:

\[
f = \frac{1}{2} \left( \frac{1}{2} x_0 + (4 + x_0) y_0 \right)^2
\]

\( -2 \leq y \leq 2 \)

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwiwel's Algorithm</th>
<th>NAG E04VDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coef ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coef ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>na</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>y_0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

\[ |d| = 1.656920 \text{ e-09} \]

\[ \Psi = -1.372127 \text{ e-08} \]

\[ \Phi(x_{*}) = 6.000000 \text{ e+00} \]

\[ f_{y=[2, 2]} = 6.000000 \text{ e+00} \]

\[ f_{y=[2,-2]} = 6.000000 \text{ e+00} \]

\[ f_{y=[-2, 2]} = 6.000000 \text{ e+00} \]

\[ f_{y=[-2,-2]} = 6.000000 \text{ e+00} \]

\[ x_{*} = -4.000000 \text{ e+00} \]

\[ y_{*} = -5.030643 \text{ e-01} \]

\[ k_{\alpha}|_{\alpha_k = 1; \forall k > k_{\alpha}} = 12 \]

| No. of iterations | 16 | 16 | 5 | 7 |
| Time | 0.2 sec | 0.2 sec | 0.2 sec | 0.1 sec |

89
Problem 12

\[ x \in \mathbb{R}^2, \; y \in \mathbb{R}^2 \]

Objective function:

\[ f = \frac{1}{2} \left( (2x_1 + 4y_0 + x_0y_0)^2 + (x_0 + 2x_0y_0 + x_1y_1)^2 \right) \]

\[-5 \leq y_i \leq 5, \; i = 0,1\]

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th>Kiwiel's Algorithm</th>
<th>NAG E04VDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-07</td>
<td>1.000000 e-07</td>
<td>1.000000 e-07</td>
<td>1.000000 e-07</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>na</td>
</tr>
<tr>
<td>Line coef c</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>na</td>
</tr>
<tr>
<td>Penalty coef C</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( x_0 = \begin{bmatrix} 1.0 \\ 2.0 \end{bmatrix} \)

\( y_0 = \begin{bmatrix} 5.0 \\ 5.0 \end{bmatrix} \)

\( ldl = 4.431337 \text{ e-07} \)

\( \Psi = -1.325589 \text{ e-10} \)

\( \Phi(x_*) = 1.657534 \text{ e+02} \)

\( f_y = [5, 5] \)

\( f_y = [-5, 5] \)

\( x* = \begin{bmatrix} -6.849316 \text{ e-01} \\ -5.411301 \text{ e-07} \end{bmatrix} \)

\( y* = \begin{bmatrix} 5.000000 \text{ e+00} \\ 5.000000 \text{ e+00} \end{bmatrix} \)

\( k_{xk} | \alpha_k = 1; \forall k > k_{x_k} = 60 \)

No. of iterations: 61

Time: 0.7 sec
Problem 13

\[ x \in \mathbb{R}^4, \ y \in \mathbb{R}^2 \]

Objective function:

\[
f = \frac{1}{2} (x_0 x_1 - y_0 (1 - x_2 x_3))^2 + (x_1 x_2 - y_1 (2 + x_3 x_0))^2 + y_0^2 + y_1^2
\]

\[-5 \leq y_i \leq 5, \ i = 0, 1\]

<table>
<thead>
<tr>
<th>Proposed Algorithm</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th>Kiwiels’ Algorithm</th>
<th>NAG E04VDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>na</td>
</tr>
<tr>
<td>Line coef c</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>na</td>
</tr>
<tr>
<td>Penalty coef C</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>na</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( x_0 )</th>
<th>( y_0 )</th>
<th>( \Phi(x^*) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.0</td>
<td>3.0</td>
<td>2.500000 e+01</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>3.0</td>
<td>2.500000 e+01</td>
</tr>
<tr>
<td></td>
<td>-1.0</td>
<td>3.0</td>
<td>2.500000 e+01</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>3.0</td>
<td>2.500000 e+01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( f_y=[ 5, 5] )</th>
<th>( f_y=[-5, 5] )</th>
<th>( f_y=[-5, -5] )</th>
<th>( f_y=[5, -5] )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.500000 e+01</td>
<td>2.500000 e+01</td>
<td>2.500000 e+01</td>
<td>2.500000 e+01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( x^* )</th>
<th>( y^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.554780 e+00</td>
<td>5.000000 e+00</td>
</tr>
<tr>
<td></td>
<td>3.240998 e-05</td>
<td>5.000000 e+00</td>
</tr>
<tr>
<td></td>
<td>7.77394 e-01</td>
<td>5.000000 e+00</td>
</tr>
<tr>
<td></td>
<td>1.286372 e+00</td>
<td>5.000000 e+00</td>
</tr>
</tbody>
</table>

| | \( k_{\alpha} | \alpha_k = 1; \forall k > k_{\alpha} \) |
|---|---|
| | 29 | 0 | 0 | na |

<table>
<thead>
<tr>
<th></th>
<th>No. of iterations</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>0.9 sec</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>0.7 sec</td>
</tr>
<tr>
<td></td>
<td>974</td>
<td>9.2 sec</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.1 sec</td>
</tr>
</tbody>
</table>

91
Problem 14

$$x \in \mathbb{R}^4, \; y \in \mathbb{R}^3$$

Objective function:

$$f = y_0^2 + y_0(x_0^2 - x_1 + x_2 - x_3 + 2) + y_1^2 + y_1(-x_0 + 2x_1^2 - x_2^2 + 2x_3 - 10) + y_2^2 + y_2(2x_0 - x_1 + 2x_2 - x_3^2 - 5) + 5x_0^2 + 5x_1^2 + 5x_2^2 + 5x_3^2$$

$$-2 \leq y_i \leq 2, \; i = 0, 1, 2$$

Proposed Algorithm

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon $\epsilon$</td>
<td>$1.000000 ; e^{-14}$</td>
<td>$1.000000 ; e^{-14}$</td>
</tr>
<tr>
<td>Lambda $\lambda$</td>
<td>$5.000000 ; e^{-01}$</td>
<td>$5.000000 ; e^{-01}$</td>
</tr>
<tr>
<td>Line coef $c$</td>
<td>$1.000000 ; e^{-04}$</td>
<td>$1.000000 ; e^{-04}$</td>
</tr>
<tr>
<td>Penalty coef $C$</td>
<td>$1.000000 ; e+06$</td>
<td>$1.000000 ; e+06$</td>
</tr>
</tbody>
</table>

| $x_0$ | $(10,1,1,10)^T$ | $(10,1,1,10)^T$ | $(10,1,1,10)^T$ | $(10,1,1,10)^T$ |
| $y_0$ | $(2,2,2)^T$ | $(2,2,2)^T$ | $(2,2,2)^T$ | $(2,2,2)^T$ |

| x* | 1.428571 e-01 | 1.428571 e-01 | 1.428572 e-01 | 1.428571 e-01 |
| y* | 2 | 2 | 2 | na |

| $k_{\alpha}$ | $1; \forall k > k_{\alpha}$ | 0 | 0 | na |

| No. of iterations | 43 | 42 | 73 | 23 |
| Time | 1.2 sec | 0.9 sec | 1.0 sec | 0.2 sec |
Problem 15

\( x \in \mathbb{R}^4 \), \( y \in \mathbb{R}^4 \)

Objective function:

\[
\begin{align*}
  f &= y_0^2 + y_0(x_0^2 - x_1 + x_2 - x_3 + 2) + y_1^2 + y_1(-x_0 + 2x_1^2 - x_2^2 + 2x_3 - 10) \\
  &\quad + y_2^2 + y_2(2x_0 - x_1 + 2x_2 - x_3 - 5) + y_3^2 + y_3(5x_0^2 + 4x_1^2) + 3x_2^2 + 2x_3^2 \\
  &\quad - 2 \leq y_i \leq 2, \quad i = 1, 2, 3, 4
\end{align*}
\]

Proposed Algorithm

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th>Kiwiel's Algorithm</th>
<th>NAG E04VDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-14</td>
<td>1.000000 e-14</td>
<td>1.000000 e-14</td>
<td>1.000000 e-08</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>na</td>
</tr>
<tr>
<td>Line coef ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>na</td>
</tr>
<tr>
<td>Penalty coef ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>1.000000 e-08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_0 )</td>
<td>((-1,-1,-1,-1)^T)</td>
<td>((-1,-1,-1,-1)^T)</td>
</tr>
<tr>
<td>( y_0 )</td>
<td>((2,2,2,2)^T)</td>
<td>((2,2,2,2)^T)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idl</td>
<td>1.908190 e-08</td>
<td>1.929356 e-08</td>
</tr>
<tr>
<td>( \Psi )</td>
<td>-7.105427 e-15</td>
<td>-2.501110 e-12</td>
</tr>
<tr>
<td>( \Phi(x_*) )</td>
<td>4.848810 e+01</td>
<td>4.848810 e+01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_\ast )</td>
<td>([-2,-2,-2,-2]^T)</td>
<td>([-2,-2,-2,-2]^T)</td>
</tr>
<tr>
<td>( k )</td>
<td>((2,2,2,2)^T)</td>
<td>((2,2,2,2)^T)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of iterations</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Time</td>
<td>0.5 sec</td>
<td>0.5 sec</td>
</tr>
</tbody>
</table>

93
2.4 Discrete Minimax Problems

Test Problems 16-20 are published discrete minimax problems with constraints in the $x$-space. Using penalty functions, we reformulate them as minimax problems unconstrained in the $x$-space.

Discrete minimax formulation:

$$\min_{x \in \mathbb{R}^n} \max_{t=1,\ldots,T} \{f_t(x)\}, \quad t = 1, \ldots, T \tag{2.5}$$

subject to:

$$g_l(x) = 0, \quad l = 1, \ldots, L$$

The continuous minimax reformulation of (2.5) is given below.

Continuous minimax formulation:

$$\min_{x \in \mathbb{R}^n} \max_{y \in Y_t} \left\{ \sum_{i=1}^{L} y_i f_i + \sum_{i=1}^{L} c g_i^2 \right\} \tag{2.6}$$

where

$$Y_t = \{y, \ 0 \leq y_i \leq 1, \sum y_i = 1\}$$

and where the constraints on $x$ are reformulated as penalty terms with $c \geq 0$.

Problem (2.5) was solved using the proposed algorithm, Kiwiel's[19] algorithm and NAG E04VDF nonlinear programming subroutine. For Test Problems 16-20 we report the values of the functions $f_t$ at the solution.
Test Problems 16-20 are published problems. We summarize the published optimal values below.

Table 2.4  Published optimal values for Test Problems 16-20.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Source</th>
<th>Optimal Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Charalambous and Bandler(^6)</td>
<td>1.95222</td>
</tr>
<tr>
<td>17</td>
<td>Charalambous and Bandler(^7)</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>Conn(^8)</td>
<td>2.25</td>
</tr>
<tr>
<td>19</td>
<td>Demyanov and Malozemov(^9)</td>
<td>-3</td>
</tr>
<tr>
<td>20</td>
<td>Polak, Mayne and Higgins(^{10})</td>
<td>e = 2.718281</td>
</tr>
</tbody>
</table>

We note that Test Problem 20 reported here, like its more general formulation found in Polak, Mayne and Higgins[32], is analytically a nonproblem because the solution is obvious by visual inspection of the objective function. The optimal value of the objective function is purely dependent on the constant coefficients and for any given constant coefficient, the optimal \(x\) is always zero.

\(^{6}\)Reported in Charalambous and Conn[6].
\(^{7}\)Reported in Charalambous and Conn[6].
\(^{8}\)Reported in Rustem[41].
\(^{9}\)Reported in Di Pillo and Grippo[12].
\(^{10}\)Reported in Di Pillo and Grippo[12].
Problem 16

Discrete minimax problem:
(Charalambous and Bandler, 1976)

\[ x \in \mathbb{R}^2 \]

\{ \begin{align*}
f_1 &= x_0^2 + x_1^4, \\
f_2 &= (2 - x_0)^2 + (2 - x_1)^2, \\
f_3 &= 2e^{(x_1 - x_0)}
\end{align*} \}

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwiell's Algorithm</th>
<th>NAG E04VDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-09</td>
<td>1.000000 e-09</td>
<td>1.000000 e-09</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>na</td>
</tr>
<tr>
<td>Line coef ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>na</td>
</tr>
<tr>
<td>Penalty coef ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>na</td>
</tr>
</tbody>
</table>

| \( x_0 \) | 1.0 | 1.0 | 1.0 |
| \(-0.1\) | -0.1 | -0.1 | na |

| \( y_0 \) | 3.333333 e-01 | 3.333333 e-01 | na |
| | 3.333333 e-01 | 3.333333 e-01 | na |
| | 3.333333 e-01 | 3.333333 e-01 | na |

| \( |d| \) | 8.711166 e-06 | 8.711166 e-06 | 3.347541 e+00 |
| \( \Psi \) | -5.762465 e-08 | -5.762465 e-08 | -5.233548 e-02 |
| \( \Phi(x^*) \) | 1.952225 e+00 | 1.952225 e+00 | 1.972011 e+00 |

| \( f_1 \) | 1.952225 e+00 | 1.952225 e+00 | 1.972011 e+00 |
| \( f_2 \) | 1.952225 e+00 | 1.952225 e+00 | 1.972011 e+00 |
| \( f_3 \) | 1.574604 e+00 | 1.574604 e+00 | 1.347779 e+00 |

| \( x^* \) | 1.138850 e+00 | 1.138850 e+00 | 1.224174 e+00 |
| | 8.997067 e-01 | 8.997067 e-01 | 8.294855 e-01 |
| \( y^* \) | 0.000000 e+00 | 0.000000 e+00 | 1.000000 e+00 |
| | 1.000000 e+00 | 1.000000 e+00 | na |
| | 0.000000 e+00 | 0.000000 e+00 | na |

| \( k_\alpha \{ k \} \) | 1; \( \forall k > k_\alpha \) | 0 | 0 | na |

| No. of iterations | 26 | 26 | 21 | 7 |
| Time             | 0.8 sec | 0.6 sec | 1.3 sec | 0.1 sec |

Remark: FAILURE: stepsize=0
Problem 17

Discrete minimax problem:
(Charalambous and Bandler, 1976)

\[ x \in \mathbb{R}^2 \]

\[ \{ f_1 = x_0^4 + x_1^2, f_2 = (2 - x_0)^2 + (2 - x_1)^2, f_3 = 2e^{(x_1 - x_0)} \} \]

<table>
<thead>
<tr>
<th>Proposed Algorithm</th>
<th></th>
<th>Kiwiels Algorithm</th>
<th>NAG E04VDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parameters:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-07</td>
<td>1.000000 e-07</td>
<td>1.000000 e-07</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coef c</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coef C</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
<tr>
<td><strong>x_0</strong></td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.1</td>
</tr>
<tr>
<td><strong>y_0</strong></td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
</tr>
<tr>
<td></td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
</tr>
<tr>
<td></td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
</tr>
<tr>
<td><strong>Idl</strong></td>
<td>3.112530 e-08</td>
<td>3.112530 e-08</td>
<td>5.565321 e+00</td>
</tr>
<tr>
<td><strong>Ψ</strong></td>
<td>-6.225166 e-08</td>
<td>-6.225166 e-08</td>
<td>-8.643960 e+00</td>
</tr>
<tr>
<td><em><em>Φ(x</em>)</em>*</td>
<td>2.000000 e+00</td>
<td>2.000000 e+00</td>
<td>2.160986 e+00</td>
</tr>
<tr>
<td><strong>f_1</strong></td>
<td>2.000000 e+00</td>
<td>2.000000 e+00</td>
<td>2.160986 e+00</td>
</tr>
<tr>
<td><strong>f_2</strong></td>
<td>2.000000 e+00</td>
<td>2.000000 e+00</td>
<td>2.160986 e+00</td>
</tr>
<tr>
<td><strong>f_3</strong></td>
<td>2.000000 e+00</td>
<td>2.000000 e+00</td>
<td>1.541817 e+00</td>
</tr>
<tr>
<td><strong>x</strong></td>
<td>1.000000 e+00</td>
<td>1.000000 e+00</td>
<td>1.098798 e+00</td>
</tr>
<tr>
<td></td>
<td>1.000000 e+00</td>
<td>1.000000 e+00</td>
<td>8.386127 e-01</td>
</tr>
<tr>
<td><strong>y</strong></td>
<td>1.000000 e+00</td>
<td>1.000000 e+00</td>
<td>1.000000 e+00</td>
</tr>
<tr>
<td></td>
<td>0.000000 e+00</td>
<td>0.000000 e+00</td>
<td>0.000000 e+00</td>
</tr>
<tr>
<td></td>
<td>0.000000 e+00</td>
<td>0.000000 e+00</td>
<td>0.000000 e+00</td>
</tr>
<tr>
<td><strong>k_α</strong></td>
<td>1; ( \forall k &gt; k_α )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>No. of iterations</strong></td>
<td>81</td>
<td>81</td>
<td>35</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>1.5 sec</td>
<td>1.4 sec</td>
<td>2.3 sec</td>
</tr>
<tr>
<td><strong>Remark</strong></td>
<td>FAILURE: stepsize=0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Problem 18

Discrete minimax problem:
(Conn, 1979)

\[ x \in \mathbb{R}^2 \]

\[ f_1 = x_0^2 + x_1^2, \quad f_2 = (2 - x_0)^2 + (2 - x_1)^2, \quad f_3 = 2e^{(x_1 - x_0)} \]

subject to:

\[ g_1 = x_0 + x_1 - 2 \]

\[ g_2 = -x_0^2 - x_1^2 + 2.25 \]

<table>
<thead>
<tr>
<th>Proposed Algorithm</th>
<th>Kiwiell's Algorithm</th>
<th>NAG E04VDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
<td></td>
</tr>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-09</td>
<td>1.000000 e-09</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coef c</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coef C</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( x_0 )</th>
<th>2.1</th>
<th>2.1</th>
<th>2.1</th>
<th>2.1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( y_0 )</th>
<th>3.333333 e-01</th>
<th>3.333333 e-01</th>
<th>3.333333 e-01</th>
<th>na</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>na</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( ld )</th>
<th>1.159692 e-06</th>
<th>1.159692 e-06</th>
<th>8.761855 e-05</th>
<th>na</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Psi )</td>
<td>-4.259504 e-10</td>
<td>-4.259504 e-10</td>
<td>-9.047239 e-10</td>
<td>na</td>
</tr>
<tr>
<td>( \Phi(x^*) )</td>
<td>2.247500 e+00</td>
<td>2.247500 e+00</td>
<td>2.247500 e+00</td>
<td>2.250000 e+00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( f_1 )</th>
<th>2.244998 e+00</th>
<th>2.244998 e+00</th>
<th>2.245000 e+00</th>
<th>2.250000 e+00</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_2 )</td>
<td>2.244998 e+00</td>
<td>2.245000 e+00</td>
<td>2.250000 e+00</td>
<td>2.250000 e+00</td>
</tr>
<tr>
<td>( f_3 )</td>
<td>9.931734 e-01</td>
<td>9.931708 e-01</td>
<td>9.861377 e+00</td>
<td>9.861377 e+00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( x^* )</th>
<th>1.349999 e+00</th>
<th>1.349999 e+00</th>
<th>1.350000 e+00</th>
<th>1.353553 e+00</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y^* )</td>
<td>1.000000 e+00</td>
<td>1.000000 e+00</td>
<td>1.000000 e+00</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>0.000000 e+00</td>
<td>0.000000 e+00</td>
<td>0.000000 e+00</td>
<td>na</td>
</tr>
</tbody>
</table>

\[ k_{\alpha_k} \mid \alpha_k = 1; \forall k > k_{\alpha_k} \]

<table>
<thead>
<tr>
<th>k_{\alpha_k}</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>na</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of iterations</td>
<td>25</td>
<td>25</td>
<td>319</td>
<td>10</td>
</tr>
<tr>
<td>Time</td>
<td>0.7 sec</td>
<td>0.8 sec</td>
<td>18.8 sec</td>
<td>0.2 sec</td>
</tr>
</tbody>
</table>

98
Problem 19

Discrete minimax problem:
(Demyanov and Malozemov, 1974)

\[ x \in \mathbb{R}^2 \]

\[ \{ f_1 = 5x_0 + x_1 \, , \, f_2 = -5x_0 + x_1 \, , \, f_3 = x_0^2 + x_1^2 + 4x_1 \} \]

**Proposed Algorithm**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th>Kiwiel’s Algorithm</th>
<th>NAG E04VDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
<td>1.000000 e-07</td>
<td>1.000000 e-07</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td></td>
</tr>
<tr>
<td>Line coef c</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td></td>
</tr>
<tr>
<td>Penalty coef C</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td></td>
</tr>
<tr>
<td>( x_0 )</td>
<td>1.05</td>
<td>1.05</td>
<td>1.10</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>1.05</td>
<td>1.10</td>
<td>1.10</td>
</tr>
<tr>
<td>( y_0 )</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td></td>
</tr>
<tr>
<td>( |d| )</td>
<td>9.200031 e-15</td>
<td>9.200031 e-15</td>
<td>2.0000007 e+00</td>
<td>na</td>
</tr>
<tr>
<td>( \Psi )</td>
<td>-3.000000 e-06</td>
<td>-3.000000 e-06</td>
<td>-2.600000 e+01</td>
<td>na</td>
</tr>
<tr>
<td>( \Phi(x^*) )</td>
<td>-2.999997 e+00</td>
<td>-2.999997 e+00</td>
<td>-2.999992 e+00</td>
<td>-3.000000 e+00</td>
</tr>
<tr>
<td>( f_1 )</td>
<td>-3.000000 e+00</td>
<td>-3.000000 e+00</td>
<td>-3.000015 e+00</td>
<td>-3.000000 e+00</td>
</tr>
<tr>
<td>( f_2 )</td>
<td>-3.000000 e+00</td>
<td>-3.000000 e+00</td>
<td>-2.999993 e+00</td>
<td>-3.000000 e+00</td>
</tr>
<tr>
<td>( f_3 )</td>
<td>-3.000000 e+00</td>
<td>-3.000000 e+00</td>
<td>-2.999993 e+00</td>
<td>-3.000000 e+00</td>
</tr>
<tr>
<td>( x^* )</td>
<td>-4.852816 e-08</td>
<td>-4.852816 e-08</td>
<td>-2.210748 e-06</td>
<td>-3.938323 e-17</td>
</tr>
<tr>
<td></td>
<td>-3.000000 e+00</td>
<td>-3.000000 e+00</td>
<td>-3.000004 e+00</td>
<td>-3.000000 e+00</td>
</tr>
<tr>
<td>( y^* )</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td>3.333333 e-01</td>
<td></td>
</tr>
<tr>
<td>( k_{\alpha_k} \alpha_k = 1 , \forall k &gt; k_{\alpha} )</td>
<td>21</td>
<td>21</td>
<td>0</td>
<td>na</td>
</tr>
<tr>
<td>No. of iterations</td>
<td>22</td>
<td>22</td>
<td>2035</td>
<td>7</td>
</tr>
<tr>
<td>Time</td>
<td>0.6 sec</td>
<td>0.6 sec</td>
<td>180.5 sec</td>
<td>0.1 sec</td>
</tr>
<tr>
<td>Remark</td>
<td>FAILURE: stepsize=0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Problem 20

Discrete minimax problem:
(Polak, Mayne and Higgins, 1991)

\[ x \in \mathbb{R}^2 \]

\[ f_1 = \exp\left(\frac{x_0}{1000} + (x_1 - 1)^2\right), \quad f_2 = \exp\left(\frac{x_0}{1000} + (x_1 + 1)^2\right) \]

---

### Proposed Algorithm

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th>Kiwiel's Algorithm</th>
<th>NAG E04VDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-14</td>
<td>1.000000 e-14</td>
<td>1.000000 e-14</td>
<td>1.000000 e-14</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>na</td>
</tr>
<tr>
<td>Line coef ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>na</td>
</tr>
<tr>
<td>Penalty coef ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>na</td>
</tr>
</tbody>
</table>

| \( x_0 \) | 2.01 | 2.01 | 2.01 | 2.01 |
| \( y_0 \) | 0.01 | 0.01 | 0.01 | 0.01 |
| \( \Psi \) | 0.5 | 0.5 | 0.5 | na |
| \( \Phi(x^*) \) | 2.718282 e+00 | 2.718282 e+00 | 2.729289 e+00 | 2.718282 e+00 |
| \( f_1 \) | 2.718282 e+00 | 2.718282 e+00 | 2.729285 e+00 | 2.718282 e+00 |
| \( f_2 \) | 2.718282 e+00 | 2.718282 e+00 | 2.729287 e+00 | 2.718282 e+00 |

| \( x^* \) | -3.924384 e-08 | -3.924384 e-08 | 2.009958 e+00 | -9.292816 e-10 |
| \( y^* \) | -4.875688 e-08 | -4.875688 e-08 | -1.817295 e-07 | 8.327223 e-18 |
| \( \Psi^* \) | 0.000000 e+00 | 0.000000 e+00 | 0.000000 e+00 | na |
| \( \Psi^* \) | 1.000000 e+00 | 1.000000 e+00 | 1.000000 e+00 | na |

| \( k_{\max} \) | 322 | 322 | 0 | na |
| No. of iterations | 323 | 323 | 14 | 8 |
| Time | 3.7 sec | 3.7 sec | 0.9 sec | 0.2 sec |

Remark: FAILURE: stepsize=0
3 Conclusion

We used a number of test problems to ascertain the performance of the proposed quasi-Newton algorithm. For the convex-concave problems, the proposed quasi-Newton algorithm performed similarly to Kiwiel's[19] algorithm; the first-order property of Kiwiel's[19] algorithm does not appear to be a disadvantage.

For the convex-convex problems, however, the quasi-Newton algorithm performed better than Kiwiel's[19] algorithm, but worse than NAG E04VDF. We suggest two reasons for the latter. The first reason is that NAG E04VDF is a nonlinear programming algorithm, applied to a nonlinear programming formulation of a discrete minimax problem. In cases where the optimum lies on the possible corner solutions, the quasi-Newton algorithm searches all possible alternatives in order to evaluate the global optimum. This may account for the time taken to converge. The second reason is that the Hessian approximation used by the quasi-Newton algorithm may become inaccurate if the estimate of the maximizer value changes frequently during successive iterations. In this case, the second-order information advantage of the quasi-Newton method may be devalued. For this class of problems, the inferior performance of Kiwiel's[19] algorithm is perhaps due to Kiwiel's[19] algorithm converging to the maximizing values at an early stage because of the linear nature of the objective function for these problems. For the rest of the iterations, Kiwiel's[19] algorithm performed as a first-order method, in contrast to the superlinear local behaviour of the quasi-Newton and NAG E04VDF algorithms.

For all the test problems, the proposed quasi-Newton algorithm found solutions that are consistent with those found by the other two algorithms, and in general, its performance is consistent with its superlinear convergence property.
1 Introduction

In Chapter 3, we discussed the performance of the proposed quasi-Newton algorithm in solving a number of test problems. In this chapter, we apply the proposed algorithm to solve a finance problem which is of particular current (June 1994) interest in that it is a hedging problem in the risk management of derivative securities. We define the generic minimax formulation of a hedging strategy, hereafter the **minimax hedging strategy** or **minimax**, and develop it into a number of specific strategies which we call **variants**. We then measure the performance of the variants against delta hedging, which we use as a benchmark throughout this study. Minimax differs from delta hedging in that it allows the hedger to incorporate information or his beliefs about the future level of prices.

In Sections 2 and 3 we give a general introduction to options and the Black and Scholes[1] option pricing model. In Section 2, we describe call options on a stock from the point of view of the hedger who writes\(^1\) one. In Section 3, we describe an option pricing model and a dynamic hedging strategy. In Section 4, we define the minimax hedging strategy. In Sections 5 and 6, we present and discuss results when minimax and delta hedging are used by the writer of a European call option\(^2\) which can be regarded as a

---

\(^1\)When an option contract is sold, the seller of the contract is called "the writer" of the contract and the act of selling is referred to as "writing".

\(^2\)An option which can be exercised only at maturity.
the American call option. In Section 5, we present a simulation study designed to identify the properties of the variants of minimax and to ascertain whether minimax performs best for a set of options for which it is designed to perform best. In Section 6, we present an empirical study to ascertain the performance of minimax when real data are used.

1.1 Minimax: the hedging strategy and the algorithm

The minimax hedging strategy is a risk management policy that uses the minimax algorithm as a tool for implementing such a policy. The minimax hedging strategy, just as any other hedging strategy, is a policy that reflects the hedger's desired risk profile. The term minimax differentiates it from other hedging strategies by giving some information on what type of policy it is: it is a policy based on the worst-case potential hedging error. Just as any other hedging strategy, the minimax hedging strategy is implemented over a period of time, say, a nine-month period corresponding to the life of an option, and it may involve more than one rebalancing date. At a rebalancing date, the hedge is adjusted to reflect the hedger's desired risk profile. At a rebalancing date, the minimax hedging strategy uses the minimax algorithm as a computation tool to find the worst-case potential hedging error and the corresponding solution.

We distinguish between the performance of the minimax algorithm and that of the minimax hedging strategy. The performance of the minimax algorithm is measured from the point of view of optimization. The relevant performance criteria such as convergence and rate of convergence are estimated by performance measures such as closeness to the solution and time taken to converge. The performance of the minimax hedging strategy is measured from the point of view of risk management in finance. The relevant performance criteria such as minimal risk and maximal return, respectively, are estimated by performance measures such as volatility of profit and actual profit. In

---

3 An option which can be exercised any time during the life of the option.
Section 5.6, we define the measure of performance used in the simulation and empirical studies.

1.2 Notation

The following notation will be used in this chapter:
- \( B \) call price
- \( S \) stock price
- \( X \) exercise price
- \( r \) risk-free interest rate
- \( t \) current date
- \( T \) expiration date
- \( T - t \) time to maturity
- \( \sigma \) volatility
- \( \Theta(d) \) the cumulative normal distribution function
- \( \Delta t \) hedging interval
- \( N \) the contracted number of shares of stock
- \( n \) number of shares to hold
- \( k \) the number of written call options
- \( K \) roundtrip transaction cost as proportion of transaction volume

The following subscripts will be used:
- \( 0 \) time 0, the initiation date of the contract
- \( t \) time \( t \), any time such that \( 0 < t < T \)
- \( T \) time \( T \), the expiration date

The following superscript will be used:
- \( i \) refers to stock \( i \) or option \( i \)

2 Introduction to options and the hedging problem

2.1 Stock options

A stock option, hereafter referred to as option, is a contract that entitles the holder to buy or sell a specific number of shares of a given stock, at or within a certain period of time, for an agreed price; the price which a buyer of an option pays is called the premium. The contract under which the option is bought or sold specifies:
• $B_0$, the premium, the price paid by the buyer,
• the underlying stock,
• $N$, the contracted number of shares,
• the date of the contract,
• $T$, time $T$, the expiration date,
• $X$, the exercise price, the price at which the offer to buy or sell is to be made.

The market for traded options is the option exchange. The two most widely traded options are called calls and puts: a call gives the holder the right to buy a specific number of shares; a put gives the holder the right to sell a specific number of shares. If the exercise of the option can take place only at the expiration date, it is called a European option. If the exercise can take place at any time on or before the exercise date, it is called an American option.

We focus on European call options, and the hedging problem we describe below pertains to the writing or selling of this type of option. Our analysis also applies to American call options, provided that the American call's underlying stock does not pay dividends. In this case, it is never optimal to exercise an American call before the expiration date.5

Our analysis does not apply to put options. Although there is a certain symmetry between puts and calls, it is not perfect. The potential profit or loss, respectively, from buying or selling a call is unlimited whereas that from buying or selling a put is limited. Further, when transaction costs are introduced, it has been shown that separate analyses are necessary.6

---

4In London, this is the London International Financial Futures Exchange (LIFFE).
5This result is discussed in Merton[24].
6This is illustrated by Neuhaus[25] who showed that separate analyses are necessary. In his dissertation, he developed separate option pricing models depending on whether the call option is bought or sold.
2.2 Determinants of call premium

The call premium is the price of the option determined by the market. Some of the factors that determine the call premium are the exercise price, the exercise date, and the volatility of returns on the underlying stock. In Section 3.2 we discuss the Black and Scholes\cite{1} option pricing model and the factors that affect the option price within the framework of their model. In the Black and Scholes world, the option price will be higher when:

- $X$, the exercise price is lower: this is because the lower the $X$, the higher the probability the call will be exercised.
- $T$, the time to maturity is longer: this is because uncertainty is greater the longer the time between the contract date and the expiration date and, for calls, there is a higher probability of a big reward.
- $\sigma$, the volatility of returns on the stock is higher: this is because high volatility increases the uncertainty about the final level of the stock price, $S_T$, and therefore increases the probability that the final stock price will exceed the exercise price.

2.3 The hedging problem: hedging the risk of writing a European call option

A writer of a European call option receives the call premium but incurs a potential liability in case of exercise by the buyer at the expiration date. The potential liability is unlimited; this is shown by a graph of profit or loss against final stock price in Figure 2.3.

\textbf{Figure 2.3} Profit or loss graph at the expiration date.
The writer, hereafter also called the hedger, wishes to modify his exposure to risk: he would like to avoid a potentially large loss in case the final stock price is above the exercise price. He would hedge this risk by holding part or all of the contracted number of shares. If he chooses to hold all of the contracted number of shares at the time the contract was made, then he has implemented a covered write strategy. This is a static hedging strategy where a decision is made at one point in time only. If he chooses to hold part of the contracted number of shares and, in particular, adjusts his holding based on the option's "delta", then he is implementing a delta hedging strategy. This is a dynamic hedging strategy where a decision is made at several points in time.

In practice, hedgers can choose from a variety of strategies ranging from ad hoc strategies to sophisticated ones based on option pricing theories. As hedging strategies with theoretical foundations are widely studied and generally acknowledged to be efficient, we assume in general that hedgers use them. In particular, we assume they use delta hedging.

We shall address this hedging problem in Section 4 where we develop the minimax hedging strategy: this strategy is based on the notion of a "minimax hedging error".

3 The Black and Scholes option pricing model and delta hedging

In a dynamic strategy the hedger modifies his position in response to movements in the stock price. In Section 3.1, we present the Black and Scholes[1] (BS) option pricing model, which is the basis of delta hedging. The BS model will also be used in the

---

7The "delta" of an option is the marginal change in the value of the option for a marginal change in the underlying stock's price. This is discussed in more detail in Section 3.
8Cox and Rubinstein[10] discuss the theoretical foundations of option valuation and present practical implementations of hedging strategies based on them. Chen and Johnson[7] also discuss strategies for hedging options.
9The minimax hedging error is defined and discussed in Section 4.5.
minimax hedging strategy to be described in Section 4. In Section 3.2, we present a method of adjusting the BS option pricing model when transaction costs are included in the option valuation. In Section 3.3, we discuss delta hedging.

3.1 Black and Scholes Option Pricing Model

Black and Scholes derived a formula for the value, $B$, of a European call option. In their model, they make $B$ a function of $S$, $X$, $\sigma$, $T-t$ and $r$, and make the following assumptions:

1. The short term interest rate is known and is constant throughout time.
2. The stock price follows a random walk in continuous time with a variance rate proportional to the square of the stock price. The distribution of possible stock prices at the end of any finite interval is lognormal. The variance rate of the return on the stock is constant.
3. The stock pays no dividend or other distributions.
4. There are no transaction costs in buying or selling the stock or option.
5. It is possible to borrow any fraction of the price of a stock to buy it or to hold it, at the short term interest rate.
6. There are no penalties to short selling.

In the Black and Scholes world, the writer can create a hedged position consisting of a position in the option and a position in the stock that is riskless for an infinitesimally short period of time. At any time $t$, he can set up a riskless portfolio because $S$, and $B$, are affected by the same underlying source of uncertainty and so are instantaneously perfectly correlated. This riskless portfolio can also be self-financing, which implies that it is not necessary to introduce, or take out, cash from the system to maintain its riskless nature. He can create a self-financing portfolio by financing any necessary purchases of shares of

\[ \text{Other option pricing models (See Cox, Ross and Rubinstein[9]) may be used but this possibility is not explored here.} \]
stock by sales of options. This can be seen as follows: let $V_H$ be the value of a "hedge"\textsuperscript{11} portfolio consisting of $n$ shares of stock held long\textsuperscript{12} and $N$ written call options,

$$V_H = nS - NB.$$  \hfill (3.1)

The dynamics of the stock price can be described by the following stochastic differential equation:

$$dS = \mu_S dt + \sigma_S S dz$$ \hfill (3.2)

where

- $\mu_S$ is the instantaneous rate of return on the stock
- $\sigma_S$ is the volatility of the rate of return on the stock
- $dt$ is an increment of time
- $dz$ is an increment of Brownian motion.

Assuming $n = n(S,t)$ and $N = N(S,t)$, using Ito's formula, the change in the value of the portfolio is

$$dV_H = dnS + ndS + dndS - [dNB + NdB + dNb]$$ \hfill (3.3)

or

$$dV_H = ndS - NdB + [dn][S + dS] - [dN][B + dB].$$ \hfill (3.4)

If the hedge portfolio is self-financing, then the sum of the third and fourth terms on the right hand side of (3.4) equals zero, i.e. $[dn][S + dS] - [dN][B + dB] = 0$ because all purchases or sales of assets are made at "new end of period prices". We now show that such a portfolio can also be riskless.

\textsuperscript{11}We use this definition of hedge portfolio in Chapter 6.

\textsuperscript{12}The purchaser of shares of stock is said to be long in the stock.
The dynamics of the option price can be described by a stochastic differential equation similar to (3.2) given by

\[ dB = \mu_B B dt + \sigma_B B dz \]  

(3.5)

where

- \( \mu_B \) is the drift rate
- \( \sigma_B \) is the volatility of the rate of the return on the option

Because of the dependency of the option price on the stock price, (3.5) can be expressed as

\[ dB = \left[ \mu_S S \frac{\partial B}{\partial S} + \frac{\partial B}{\partial t} + \frac{1}{2} \sigma_S^2 S^2 \frac{\partial^2 B}{\partial S^2} \right] dt + \sigma_S S \frac{\partial B}{\partial S} dz \]  

(3.6)

Eqn (3.4) becomes:

\[ dV_{II} = n \left[ \mu_S S dt + \sigma_S S dz \right] - N \left[ \mu_B B dt + \sigma_B B dz \right] \]  

(3.7)

In order to construct a riskless portfolio, the \( dz \) terms in Eqn (3.7) must cancel out: \( n \) should be chosen such that

\[ n \sigma_S S = N \sigma_B B \]  

(3.8)

or

\[ n \sigma_S S = N \sigma_S S \frac{\partial B}{\partial S} \]  

(3.9)

or

\[ n = N \frac{\partial B}{\partial S} \]  

(3.10)

Therefore, if the writer has \( N \) call options, he should choose \( N \frac{\partial B}{\partial S} \) shares to immunize the risk of writing the calls. In other words, if he sells \( N \) call options, he should purchase \( N \frac{\partial B}{\partial S} \) shares to make the portfolio riskless. However, the portfolio is riskless.
only for an infinitesimally short period of time; in order to maintain a riskless portfolio, the writer has to rebalance continuously. \( \frac{dP}{dS} \) is called the delta and the strategy for maintaining a riskless portfolio using it is called delta hedging. We discuss delta hedging further in Section 3.3.

The above analysis only assumes non-satiation, i.e. investors prefer more wealth to less, and that the option is neither a dominant\(^{13}\) nor a dominated security. Explicit assumptions about equilibrium or about investors' preferences are not necessary. The fundamental requirement is the non-existence of arbitrage opportunities. In this sense, the above result is preference independent because all assets are perfect substitutes for each other instantaneously and the strategy for maintaining a riskless portfolio is independent of the hedger's attitude to risk.

The Black and Scholes Formula:

\[
B = S\Theta(d_1) - Xe^{-(T-t)}\Theta(d_2) \quad (3.11)
\]

\[
d_1 = \frac{\ln(S/K) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}} \quad (3.12)
\]

\[
d_2 = \frac{\ln(S/K) + (r - \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}} = d_1 - \sigma\sqrt{T-t} \quad (3.13)
\]

where

\[
\Theta(d) \text{ the cumulative normal distribution function, i.e.} \\
\Theta(d) = \int_{-\infty}^{d} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad (3.14)
\]

\(^{13}\text{From Merton}[24], \text{ Security A is dominant over Security B if, on some known date in the future, the return on A will exceed the return on B for some possible states of the world, and will be at least as large as on B, in all possible states of the world.}\]
3.2 The Leland Model: introduction of transaction costs

The BS option pricing model does not include transaction costs. Leland[20] developed an option pricing model that includes transaction costs. In his model, the hedging errors, including transaction costs, will almost surely approach zero as $\Delta t \to 0^{14}$. His model was extended by Neuhaus[25]. Leland and Neuhaus both incorporated transaction costs in their models by modifying the estimated volatility. We present below Leland's modification, which is more closely related to discrete delta hedging than is the Neuhaus modification.

The revised volatility $\hat{\sigma}$ is given by

$$\hat{\sigma} = \sqrt{\sigma^2 \left[1 + \frac{2}{\pi} \frac{K}{\sigma \sqrt{\Delta t}} \right]} \quad (3.15)$$

where

$K^{15}$ is the roundtrip$^{16}$ transaction cost expressed as a proportion of trading volume.

We replace $\sigma$ by $\hat{\sigma}$ in the BS option pricing model when we include transaction costs in the analysis. We define $K = 2\hat{K}$, where $\hat{K}$ is one half the roundtrip transaction cost.

3.3 The delta of a call option and delta hedging

Delta and delta hedging are closely related to the BS option pricing model$^{17}$. Delta, $D$, is the change in option price per unit change in stock price, that is

$$D = \frac{\partial P}{\partial S} \quad (3.16)$$

---

$^{14}$Other option pricing models with transaction costs have been developed by other authors. We used Leland's[20] model because it is applicable to discrete rebalancing and it fits in the framework of the minimax hedging strategy. See Boyle and Emanuel[3], Gilster and Lee[17], Panas[26], Neuhaus[25].

$^{15}$In Leland[20], transaction costs varied from $K = 0.0$ to $K = 0.04$.

$^{16}$From Rosenberg[37], a roundtrip trade is defined as any complete transaction made up of a buy followed by a sale of the same stock or vice versa.

$^{17}$The model and its relevance to delta hedging are described in Section 3.1.
Using the BS option pricing model Eqn (3.11), delta is given analytically by

\[ D = \Theta(d_1). \]  \hspace{1cm} (3.17)

Black and Scholes argue that the writer can realize a riskless portfolio by delta hedging; he finds the delta of an option and bases the number of shares to hold on it. He sets the **hedge ratio**, defined as the number of shares that he must hold divided by the number of shares per option at any time, equal to delta. For \( N \), the contracted number of shares per option, and \( n_t \), the number of shares to hold at any time \( t \), under delta hedging,

\[ n_t = D_t N. \]  \hspace{1cm} (3.18)

Because delta changes with time, the hedge portfolio is riskless only for a very short period of time. The hedger must rebalance continuously to keep the portfolio riskless. Such a portfolio strategy is called a "**delta-neutral" strategy. However, because he cannot rebalance continuously, he uses discrete delta hedging where he rebalances at discrete intervals of time. With discrete delta hedging, he incurs a **hedging error** which, for an interval of time, is the net position of a hedge portfolio brought about by changes in \( S_t \). The hedging error (\( HE \)) for a portfolio of a written call option and stock held long, for the interval \( t \) to \( t+1 \) is:

\[ HE = N(B_t - B_{t+1}) + n_t(S_{t+1} - S_t) \]  \hspace{1cm} (3.19)

Large hedging errors tend to increase the cost of rebalancing the hedge.
4 Minimax hedging strategy

In this section, we develop a strategy to solve the hedging problem which was introduced in Section 2.3. This strategy is based on the concept of a worst case scenario, which the hedger specifies in terms of movements in stock price, and it finds a hedge that minimizes the effect of such a scenario. In Section 4.1, we formulate the minimax problem. In Section 4.2, we present two worst case scenarios. In Section 4.3, we present the hedging error which is the underlying cost to be minimized. In Section 4.4, we present the objective functions. In Section 4.5, we return to the hedging error in the context of minimax. In Section 4.6, we discuss the treatment of transaction costs. In Section 4.7, we present the variants of the minimax hedging strategy. In Section 4.8, we discuss the minimax solution to the problem given in Section 4.1.

4.1 Minimax problem formulation

The problem is to minimize an objective function under a worst case scenario. The minimizing variable is \( n \), and the maximizing variable is \( S_{t+1} \), which is allowed to take any value within predefined bounds. The minimax problem is

\[
\min_{n} \max_{S_{t+1}} f(n, S_{t+1}) \quad \text{(4.1)}
\]

subject to \( S_{t} \leq S_{t+1} \leq S_{u} \) \quad \text{(4.2)}

where \( f(n, S_{t+1}) \) is either Objective Function 1 or Objective Function 2, presented in Section 4.4, and \( S_{t} \leq S_{t+1} \leq S_{u} \) is either the range defined under Worst Case 1, presented in Section 4.2.1, or under Worst Case 2, presented in Section 4.2.2.

There are no constraints on \( n \), the number of shares to hold at time \( t \): non-negative \( n \) implies a long position in shares\(^{18}\); negative \( n \) implies a short position in shares\(^{19}\).

---

\(^{18}\)The hedger bought the shares.

\(^{19}\)The hedger sold the shares. This situation is possible in the minimax context.
4.2 The worst case scenario

As described above, at time \( t \), the minimax strategy computes the hedging error that corresponds to the worst case scenario that may occur at time \( t + 1 \), which is defined as the movements of stock price that are unfavorable to the hedger. In solving the minimax problem, we maximize over \( S_{i,t+1} \), where \( S_{i,t+1} \) is within the range

\[ S_l \leq S_{i,t+1} \leq S_u \]  \hspace{1cm} (4.3)

We now consider two worst case scenarios, given in Sections 4.2.1 and 4.2.2.

4.2.1 Worst Case 1

In Worst Case 1, the hedger defines the worst case in terms of extreme movements in stock price. The range of \( S_{i,t+1} \) has upper and lower bounds that delimit the 95% confidence interval\(^{20}\) of all possible values of the future stock price, i.e. two standard deviations about the expected value of the stock price at time \( t + 1 \). This 95% confidence interval is based on an estimate of the volatility of the stock price and on the assumption that the stock price follows a lognormal distribution function.

Worst Case 1 is hereafter referred to as the 95%-Level.

4.2.2 Worst Case 2

We define the state of the option as one of three possibilities:

- in-the-money\(^{21}\)
- at-the-money\(^{22}\)
- out-of-the-money\(^{23}\).

Most of the business in exchange traded options is in at-the-money options. The prices of the underlying stocks of these options usually oscillate about the exercise price\(^{24}\).

---

\(^{20}\)We consider the 95% confidence interval as a reasonable range to use under normal market conditions. Under abnormal market conditions, the 99% confidence interval may be more appropriate.

\(^{21}\)If the current stock price is greater than the exercise price, the option is said to be in-the-money.

\(^{22}\)If the current stock price is equal to the exercise price, the option is said to be at-the-money.

\(^{23}\)If the current stock price is lower than the exercise price, the option is said to be out-of-the-money.
moving from sometimes being in-the-money to sometimes being out-of-the-money and vice versa. In Worst Case 2, we focus on movements of the stock price which may result in a switch in the state of the option, i.e. from being in-the-money at time $t$, to being out-of-the-money at time $t+1$, and vice versa. Such a switch increases the probability of a higher hedging error being incurred from $t$ to $t+1$. We define the range of $S_{t+1}$ as the range whose upper and lower bounds delimit the possible values of the future stock price within **one and three standard deviations** from the expected value of the stock price at time $t+1$, in the direction of the exercise price $X$. This means that if $S_t > X$, the relevant range would be on the left side of the distribution of future stock price; if $S_t \leq X$, the relevant range would be on the right side.

Worst Case 2 is hereafter referred to as the **Abrupt-Change**.

**4.3 The hedging error**

From Section 3.3, under delta hedging, the hedging error is given by Eqn (3.19). When actual values of $B_t$, $S_t$, $B_{t+1}$, and $S_{t+1}$ are substituted into Eqn (3.19), we have the actual hedging error under delta hedging. When actual values of $B_t$ and $S_t$ and potential values of $B_{t+1}$ and $S_{t+1}$ are substituted into Eqn (3.19), we have the potential hedging error under delta hedging. The potential hedging error under delta hedging is the basis of the objective functions in the minimax hedging strategy. In minimax, potential $S_{t+1}$ is taken from a predefined range that maximizes the objective function; potential $B_{t+1}$ is the value of the call option based on the pricing model\(^{25}\) given potential $S_{t+1}$, i.e. potential $B_{t+1} = B_{t+1}(S_{t+1})$. The minimax strategy minimizes the maximum potential hedging error plus interest payments on borrowed money\(^{26}\). In Section 4.5, we define the minimax hedging error and give the definition of actual hedging error and potential hedging error in the context of minimax.

\(^{24}\)This is the region of greatest elasticity (curvature) and therefore the area about which market makers are mostly concerned. Formally, we are talking about the effects of the option's "gamma".

\(^{25}\)We use the Black and Scholes\(^{11}\) option pricing model, Eqn (3.11), with a modified volatility, Eqn (3.15).

\(^{26}\)This is discussed in detail in Section 4.4.3.
4.4 The objective functions

4.4.1 The objective of minimax hedging
In any dynamic hedging strategy, hedging errors are incurred; in order to correct for these errors, the hedge is rebalanced, with the cost of rebalancing being added to the cost of hedging. At time $t$, the hedger can attempt to minimize the potential hedging error between $t$ and $t+1$. His decision at time $t$ on $n_t$, the number of shares to hold, affects the actual hedging error between $t$ and $t+1$. The minimax hedging strategy aims to minimize the maximum potential hedging error between $t$ and $t+1$, indirectly through Objective Function 1, and directly through Objective Function 2.

4.4.2 Objective Function 1
The minimax strategy aims to minimize the potential hedging error indirectly by minimizing the delta of the whole portfolio (stocks plus written call options), i.e. it attempts to achieve delta neutrality. This indirect way of minimizing the potential hedging error has its theoretical basis in delta hedging. Neuhaus[25] examined the properties of delta hedging based on the BS option pricing model and reports the following:

- the expected hedging error is zero;
- the variance of the hedging errors is minimized.

Thus, a portfolio that is continuously kept delta-neutral should theoretically have a zero cumulative hedging error. In Section 3.1, it has been shown that in the BS world, a riskless portfolio could be maintained by continuous rebalancing using delta hedging. In other words, the delta is the optimal hedge ratio. The minimax strategy attempts to achieve delta neutrality at time $t$ and at time $t+1$; i.e. it minimizes the maximum deviation of the portfolio's delta at time $t$ and at time $t+1$ from zero.
4.4.2.1 Objective Function 1 defined

We define \( U_1 : \mathbb{R}^k \rightarrow \mathbb{R}^l \), \( U_2 : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^l \), \( U : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^2 \), \( n_i \in \mathbb{R}^k \), \( S_{i+1} \in \mathbb{R}^k \) and \( Q \) as a (2x2) positive definite diagonal weighting matrix, with elements \( q_1 \) and \( q_2 \). \( U^d \in \mathbb{R}^2 \) is the vector of desired values: we use a desired value of zero, i.e. the desired portfolio delta is zero or that the portfolio is delta-neutral.

The objective function is given by

\[
f(n_i, S_{i+1}) = \frac{1}{2} < U - U^d, Q(U - U^d)> \tag{4.4}
\]

where

\[
\begin{bmatrix}
    n_i \\
    \vdots \\
    n_i
\end{bmatrix} \quad \text{and} \quad 
\begin{bmatrix}
    S_{i+1} \\
    \vdots \\
    S_{i+1}
\end{bmatrix} \tag{4.5}
\]

\[
U(n_i, S_{i+1}) = 
\begin{bmatrix}
    U_1(n_i) \\
    U_2(n_i, S_{i+1})
\end{bmatrix} \quad \text{and} \quad U^d = 
\begin{bmatrix}
    U^d_1 \\
    U^d_2
\end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4.6}
\]

\[
U_1(n_i) = \sum_{i=1}^{k} n_i^i D_{S_i}^i - \sum_{i=1}^{k} N^i D_{S_{i+1}}^i \tag{4.7}
\]

\[
U_2(n_i, S_{i+1}) = \sum_{i=1}^{k} n_i^i D_{S_{i+1}}^i - \sum_{i=1}^{k} N^i D_{S_{i+1}}^i (S_{i+1}^i) \tag{4.8}
\]

where

\( D_{S_i}^i \) is the delta of stock \( i \) at time \( t \)

\( D_{S_{i+1}}^i \) is the delta of call option \( i \) at time \( t \)

\( ^{27} \)We give the general definition for \( k \) assets. In this chapter and in Chapter 5, we limit the study to single assets. In Chapter 6, we consider portfolios of \( k \) assets.
\( D^i_{S,t+1} \) is the delta of stock \( i \) at time \( t+1 \)

\( D^i_{B,t+1}(S'_{t+1}) \) is the delta of call option \( i \) at time \( t+1 \), as a function of \( S'_{t+1} \)

\( i = 1, \ldots, k \).

In the weighting matrix \( Q \), the weights \( q_1 \) and \( q_2 \) which are specified by the hedger, represent his preferences: with a high \( q_1 \) and a low \( q_2 \), he prefers delta neutrality at time \( t \); with a low \( q_1 \) and a high \( q_2 \), he prefers delta neutrality at time \( t+1 \).

Because the delta of a stock is 1, the minimax problem formulation can be simplified to:

\[
\min_{n_i} \max_{S_{t+1}} \left\{ q_1 \left( \sum_{i=1}^{r} n_i - \sum_{i=1}^{r} N^i D^i_{B,t+1}(S'_{t+1}) \right)^2 \right\} + \frac{1}{2} \left( \sum_{i=1}^{r} q_2 \left( \sum_{i=1}^{r} n_i - \sum_{i=1}^{r} N^i D^i_{B,t+1}(S'_{t+1}) \right)^2 \right)
\]

subject to \( S'_{t} \leq S'_{t+1} \leq S'_{u} \), \( i = 1, \ldots, k \) \hspace{1cm} (4.9)

From the BS option pricing model, delta is a function of the stock price, among other variables. In the minimax strategy, the objective function is maximized over a predefined range of future stock prices, and delta is assumed to be a function of \( S'_{t+1} \) only; other variables, i.e. risk-free interest rate, time to maturity, etc., are assumed to be known or preset. For each option \( i = 1, \ldots, k \), \( D^i_{B,t+1}(S'_{t+1}) \) is (omitting the superscript \( i \) for simplicity):

\[
D^i_{B,t+1}(S'_{t+1}) = \frac{\partial B_{t+1}}{\partial S_{t+1}} \hspace{1cm} (4.11)
\]

From Eqns (3.11), (3.12) and (3.13), and using Eqn (3.15),

\[
B_{t+1}(S_{t+1}) = S_{t+1} \Theta(d_1) - X e^{-r(t-t')} \Theta(d_2) \hspace{1cm} (4.12)
\]
\[ d_1 = \frac{\ln(S_{r,T}) + (r + \frac{\sigma^2}{2})(T-t-1)}{\tilde{\sigma}\sqrt{T-t-1}} \]  
(4.13)

\[ d_2 = \frac{\ln(S_{r,T}) + (r - \frac{\sigma^2}{2})(T-t-1)}{\tilde{\sigma}\sqrt{T-t-1}} \]
(4.14)

Eqn (4.11) becomes:

\[ D_{r,s+s} (S_{r,t}) = \Theta(d_1) \]  
(4.15)

with \( d_1 \) coming from Eqn (4.13).
4.4.3 Objective Function 2

The minimax strategy aims to minimize the potential hedging error directly by using it as the major part of the objective function. In discrete delta hedging, where rebalancing is done at discrete intervals, we expect that the desirable properties of delta hedging given in Section 4.4.2 will not be observed consistently in time. By minimizing the maximum potential hedging error plus interest payments on borrowed money, should the worst case occur, the hedger adopts a cautious strategy. If the worst case occurs, he has effectively minimized its worst effect; if it does not occur, he may incur a hedging error higher than that using delta hedging.

This direct way of minimizing the potential hedging error is based on the no-arbitrage argument of Merton[24] where he considered a portfolio containing an option, the underlying stock and a riskless bond (i.e. riskless in the sense of default) that is suitably chosen such that the aggregate investment in the portfolio is zero. He demonstrated that there is a strategy of finding the mix of option, stock and bond that would ensure that the return on the portfolio would be nonstochastic. Because of the condition of zero aggregate investment, he argued that in order to avoid arbitrage profits, the return on this portfolio must be zero. In the case of a portfolio of written call options, underlying stock and bonds, given Merton's assumptions, the return on this particular portfolio must be zero. We shall work with such a portfolio and we shall call it the "ideal portfolio". This "ideal portfolio" is the benchmark we used in defining Objective Function 2. We derive basic properties of the minimax hedging strategy on the basis of an ideal portfolio and add the effect of interest payments on borrowed money. We will return to this when we discuss Eqn (4.19) below.

---

28 Refer to Eqn (4.19) below.
4.4.3.1 Objective Function 2 defined

We define \( U_1 : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^k \), \( U_2 : \mathbb{R}^k \rightarrow \mathbb{R}^k \), \( U : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^{k+1} \), \( n_i \in \mathbb{R}^k \), \( S_{t+1} \in \mathbb{R}^k \) and \( Q \) as a \((k+1) \times (k+1)\) positive definite diagonal weighting matrix. \( U^d \in \mathbb{R}^{k+1} \) is the vector of desired values for the potential hedging error and the transaction cost terms: we use a desired value of zero, i.e. the desired hedging error is zero\(^{29}\) and the desired transaction cost is zero.

The objective function is given by

\[
 f(n_i, S_{t+1}) = \frac{1}{2} < U - U^d, Q(U - U^d) > 
\]

where

\[
 n_i = \begin{bmatrix} n_i^1 \\ \vdots \\ n_i^k \end{bmatrix} \quad \text{and} \quad S_{t+1} = \begin{bmatrix} S_{t+1}^1 \\ \vdots \\ S_{t+1}^k \end{bmatrix}
\]

\[
 U(n_i, S_{t+1}) = \begin{bmatrix} U_1(n_i, S_{t+1}) \\ \vdots \\ U_2(n_i) \end{bmatrix} \quad \text{and} \quad U^d = \begin{bmatrix} U_{d1} \\ \vdots \\ U_{d2} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}
\]

\[
 U_1(n_i, S_{t+1}) = \sum_{i=1}^{k} n_i^i (S_{t+1}^i - S_i^i) + \sum_{i=1}^{k} N_i^i (B_i^i - B_{t+1}^i (S_{t+1}^i)) + \sum_{i=1}^{k} \left( -(n_i^i - n_{i-1}^i) S_i^i + C_{i-1}^i (1 + \Delta t) \right) \Delta t 
\]

where

\[
 C_{i-1}^i = C_{i-2}^i (1 + \Delta t) - (n_{i-1}^i - n_{i-2}^i) S_{t-1}^i - \hat{K}_1 (n_{i-1}^i - n_{i-2}^i) S_{t-1}^i. 
\]

\[
 U_2(n_i) = \begin{bmatrix} U_2^1(n_i) \\ \vdots \\ U_2^k(n_i) \end{bmatrix}
\]

\(^{29}\)In minimax, we can adopt any desired value. We adopted a desired value of zero because in delta hedging, the expected value of the hedging error is zero.
We first identify all the variables in Eqns (4.16) to (4.22) and then give an economic interpretation of Eqn (4.19).

\( C_{t-1}^i \) is the cumulative value of cash inflow minus cash outflow at time \( t-1 \). \( C_{t-1}^i(1+r\Delta t) \) is \( C_{t-1}^i \) with interest payments. The first term of (4.20) is the cumulative value of cash inflow minus cash outflow from the previous period with interest payments. The second term is a cash outflow if the \( n_{t-1}^i > n_{t-2}^i \); otherwise, it is a cash inflow. The third term is always a cash outflow. We note that \( C_{t-1}^i \) will normally be a negative number. At time \( t \), \( C_{t-1}^i \) is a constant: all the variables in (4.20) have actual values.

Because transaction costs introduce nondifferentiability into the equation, they do not come into the objective function directly as part of \( U_1 \). Instead, we introduce a penalty term, \( U_2^{\prime} \) to represent a penalty for transaction costs for each option \( i \) at time \( t \). The treatment of transaction costs is discussed in detail Section 4.6. A simulation study of transaction costs is discussed in Appendix 2.

In the weighting matrix \( Q \), the weights \( q_i, \ i = 1,...,k+1 \), which are specified by the hedger, represent his preferences: with a high \( q_1 \) he prefers to minimize the potential hedging error that may be incurred from time \( t \) to time \( t+1 \); with a high \( q_i \), \( i = 2,...,k+1 \), he prefers to minimize the penalty term.

For each option \( i, i = 1,...,k \), \( B_{i+1}^{\prime}(S_{it}) \) is determined using Eqns (4.12), (4.13) and (4.14) in Section 4.4.2.1 using the modified volatility estimate given in Section 3.2.

We now give an economic interpretation of (4.19). \( U_1 \) represents the potential hedging error, inclusive of interest payments on borrowed money, between time \( t \) and time \( t+1 \):

\[ U_2^{\prime} (n_{it}^i) = \hat{K}(n_{it}^i - n_{it-1}^i)S_{it}^i. \]  (4.22)

\[ \text{where } U_2^{\prime} (n_{it}^i) = \hat{K}(n_{it}^i - n_{it-1}^i)S_{it}^i. \]
it comprises the potential shift in the stock position, the potential shift in the option position and the potential interest payment. The first two terms of (4.19) give the return on a portfolio of written call options and underlying stocks. The third term represents the opportunity cost of money, i.e. the interest payments on borrowed money, because the portfolio is not self-financing. We wish to find the mix of options and stocks that minimizes the deviation of the return on the portfolio, including opportunity cost, from the return on the "ideal portfolio", the value of which is zero, based on Merton's[24] conditions of zero aggregate investment and no-arbitrage.

4.5 The minimax hedging error

In minimax, we distinguish actual from potential hedging error. Actual hedging error, inclusive of interest payments on borrowed money, is calculated when actual $B_t$, $S_t$, $B_{t+1}$ and $S_{t+1}$ are substituted into Eqn (4.19). Potential hedging error, inclusive of interest payments on borrowed money, is calculated when actual values of $B_t$ and $S_t$ and potential values of $B_{t+1}$ and $S_{t+1}$ are substituted into Eqn (4.19). Potential $S_{t+1}$ is taken from a predefined range that maximizes the objective function; potential $B_{t+1}$ is the value of the call option based on the pricing model\[31\] given potential $S_{t+1}$, i.e. potential $B_{t+1} = B_{t+1}(S_{t+1})$. The minimax strategy minimizes the maximum potential hedging error, inclusive of interest payments on borrowed money.

We define the minimax hedging error at time $t$ as

$$\text{minimax hedging error} = U_t(n_{t*}, S_{t+1*}).$$

(4.23)

The minimax hedging error is the potential hedging error, inclusive of interest payments on borrowed money, given the solution $n_{t*}$ and $S_{t+1*}$.

\[31\]We use the Black and Scholes[1] option pricing model, Eqn (3.11), with a modified volatility, Eqn (3.15).
4.6 Transaction Costs

In this section, we discuss the treatment of transaction costs in the minimax hedging strategy. The roundtrip transaction cost \( K \) is used in valuing the option and \( \hat{K} \), with \( \hat{K} = \frac{1}{2} K \), is used as part of the cumulative value of cash inflow minus cash outflow and as part of the penalty term in Objective Function 2. In Section 4.6.1, we discuss transaction costs as part of the cumulative value of cash inflow minus cash outflow; in Section 4.6.2, we discuss transaction costs as part of Objective Function 2.

4.6.1 Transaction costs in the cost of hedging

The performance of delta hedging and the variants of minimax is measured by the final cumulative value of cash inflow minus cash outflow at the maturity of the option. After finding \( nt \) by solving the minimax problem using Eqns (4.16) to (4.22), we can evaluate the actual cumulative value of cash inflow minus cash outflow at time \( t \). This is given by

\[
C_{i} = C_{i-1}(1 + r\Delta t) - (n_{i}^{\prime} - n_{i-1}^{\prime})S_{i}^{\prime} - \hat{K}(n_{i}^{\prime} - n_{i-1}^{\prime})S_{i}^{\prime}\, (4.24)
\]

The last term is the transaction cost at time \( t \): this is always incurred and it is always a cash outflow. At time \( t = 0 \), the actual cumulative value of cash inflow minus cash outflow includes the option premium which is a cash inflow. This is given by

\[
C_{0} = -n_{0}^{\prime}S_{0}^{\prime} + NB_{0} - \hat{K}n_{0}^{\prime}S_{0}^{\prime}\, (4.25)
\]

All variants of the minimax hedging strategy will use Eqns (4.24) and (4.25) to compute the actual cumulative value of cash inflow minus cash outflow.

---

32See Section 3.2.
4.6.2 The transaction cost term in Objective Function 2

From Eqns (4.16), (4.21) and (4.22), the transaction cost term (TC) in Objective Function 2 can be expressed as

\[
TC = \sum_{i=1}^{k} q_i (U_{1i}^d - U_{d2}^i)^2 = \sum_{i=1}^{k} q_i (K(n_i + n_{i-1})S_i^2).
\]  

(4.26)

Eqn (4.26) should be interpreted as a penalty term. Because we chose a desired value of zero for \( U_2^d \), the right hand side equality holds. The effect of this term on the solution is dependent on the level of transaction cost \( K \) and on the weights \( q_i \). We use a uniform weighting system for this study: \( q_i = q, \ i = 1, \ldots, k \), where \( q \) is held constant.

For low values of \( K \), we need a high value of \( q \) so that the transaction cost term \( TC \) is not dominated by the other terms in the objective function. Conversely, for high values of \( K \), we need a low value of \( q \) to ensure that \( TC \) does not dominate the objective function. In Appendix 2, we present a study of the variation in the performance of minimax for different levels of the roundtrip transaction cost \( K \) and different values of \( q \).

For the simulation and empirical study of the different variants of minimax, we set\(^{33} \) \( K = 0.02 \) and \( q = 100 \). Simulation results are discussed in Section 5. Empirical results are discussed in Section 6.

\(^{33}\)This value is based on simulation results, reported in Appendix 2, showing the variation of \( K \) with \( q \).
4.7 The variants of the minimax hedging strategy

We consider delta hedging and 7 variants of minimax, given as variants A, B, C, D, D*, E and F, where a variant has a specific objective function and worst case scenario. Transaction costs are included in computing the cumulative value of cash inflow minus cash outflow for all the strategies considered.

Table 4.7. The hedging strategies to be used in the empirical and simulation studies.

<table>
<thead>
<tr>
<th>Code</th>
<th>Strategy</th>
<th>Objective Function</th>
<th>Condition of $S_{t+1}$</th>
<th>Transaction Costs in Objective Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Minimax</td>
<td>delta neutrality</td>
<td>n.a</td>
<td>No</td>
</tr>
<tr>
<td>B</td>
<td>Minimax</td>
<td>delta neutrality</td>
<td>95% Level</td>
<td>No</td>
</tr>
<tr>
<td>C</td>
<td>Minimax</td>
<td>potential hedging error</td>
<td>95% Level</td>
<td>No</td>
</tr>
<tr>
<td>D</td>
<td>Minimax</td>
<td>potential hedging error</td>
<td>Abrupt Change</td>
<td>No</td>
</tr>
<tr>
<td>D*</td>
<td>(Described in the next paragraph.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>As C</td>
<td>As C</td>
<td>As C</td>
<td>Yes</td>
</tr>
<tr>
<td>F</td>
<td>As D</td>
<td>As D</td>
<td>As D</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The version of D, D*, is a weighted version in which the minimax recommendation on the number of shares to hold is weighted by a factor ranging from 0 to 1 that represents the hedger's assessment of the information contained in changes in the price of the underlying stock.\(^{35}\)

---

\(^{34}\)Delta with gamma hedging is not considered because it is not a feasible strategy when considering single options.

\(^{35}\)This is a heuristic method of adjusting the hedge recommendations. The method is given in Appendix 1.
4.8 The minimax solution

Because the hedge recommendation under minimax is different from the hedge recommendation under delta hedging, in the Black and Scholes[1] world, the hedge recommendation under minimax is suboptimal. In the minimax hedging strategy, for any fixed $n_i$, we determine $S_{i+1}$, from the predefined range $S_i \leq S_{i+1} \leq S_u$, that maximizes the hedging error. We therefore can identify theoretically all the maxima corresponding to all the possible values of $n_i$. The strategy calculates the $n_i$ that minimizes over these maxima. Although the number of shares $n_i$ introduces some risk into the portfolio because it is not the same as the hedge recommendation under delta hedging, $n_i$ ensures that if the actual $S_{i+1}$, as opposed to the minimax value, falls within the range $S_i \leq S_{i+1} \leq S_u$, the absolute value of the actual hedging error, inclusive of interest payments on borrowed money, will not be worse (higher) than the absolute value of the minimax hedging error. This is the minimax robustness property. The $n_i$ value thus computed results in a robust strategy that is non-inferior in performance for any stock price within the predefined range.

Given the 95%-Level as the worst case scenario, $S_i \leq S_{i+1} \leq S_u$, the minimax algorithm ensures that either

- $n_i$ is chosen such that, for extreme point maximizers, the objective function value is the same for both upper and lower limits, or
- the objective function value corresponds to a worst case price that is in the middle of the range.

Given the Abrupt-Change as the worst case scenario, $S_i \leq S_{i+1} \leq S_u$, the minimax algorithm ensures that either

- $n_i$ is chosen such that, for extreme point maximizers, the objective function value for the upper limit is as close as possible to the value for the lower limit, or
- the objective function value corresponds to a worst case price that is in the middle of the range.

In all cases, it can be shown that the chosen $n_i$ places an upper bound on the absolute value of the hedging error that can be incurred for any price in the given range.
5 Simulation study of the performance of different strategies

In this section we describe the simulation of the performance\(^{36}\) of minimax variants against that of delta hedging when they are used to hedge the risk of writing a European call option. We present our method of generating time series of options and of their underlying stock, and of categorizing the options to different groups. We give the results of the simulation and of tests of hypotheses on the performance of the variants of minimax against delta hedging, between themselves and between groups.

5.1 Objective of the simulation

The objective of the simulation is to identify which variants, for which groups of options, outperform\(^{37}\) delta hedging, and so establish the characteristics of options, in terms of crossovers and abrupt changes\(^{38}\) in the price of the underlying stock, for which the different variants are particularly suited.

5.2 Generation of data

Two types of time series are needed in the simulation: time series of stock prices and time series of option prices. The time series of stock prices is generated using random numbers, described below, and the time series of option prices is generated from the time series of stock prices using the Black and Scholes\(^{[1]}\) option pricing model and some preset parameters. We describe how the time series of the price of the underlying stock were generated and how the final data set was created. Because we used the standard BS model to generate the option time series, we do not describe how we generated that series.

\(^{36}\)Defined in Section 5.6.
\(^{37}\)Defined in Section 5.6.
\(^{38}\)Crossovers and abrupt changes are defined in Section 5.3.
Step 1. Generation of one point (one stock price) in a time series

The NAG library routine G05CAF was used to generate two random numbers \( u_1, u_2 \in (0,1) \). This routine calls another NAG library routine G04CCF which provides the seed needed by routine G05CAF. Routine G04CCF uses the time clock of the computer to provide the seed. The random numbers \( u_1, u_2 \in (0,1) \) are used to generate two new random variables \( y_1, y_2 \), independently distributed according to the normal distribution, using the Box-Muller transformation:\(^{39}\)

\[
y_1 = \sqrt{-2 \ln u_1 \cos 2\pi u_2}
\]

and

\[
y_2 = \sqrt{-2 \ln u_1 \sin 2\pi u_2}.
\]

The random variable \( R = y_1 \) is used in the data generation. This random variable \( R \) follows a normal distribution with mean 0 and standard deviation 1, i.e. \( R \sim \Theta[0,1] \). \( R \) is transformed into a new random variable \( S_{t+1} \), the stock price at time \( t+1 \). The assumption of lognormality of stock prices with positive drift\(^{40}\) is used. This leads to the use of the following model:

\[
\ln S_{t+1} \approx \Theta[\tilde{\mu}, \tilde{\sigma}]
\]

where

\[
\tilde{\mu} = \ln S_t + (r - \frac{\sigma^2}{2})(\Delta t)
\]

\[
\tilde{\sigma} = \sigma \sqrt{\Delta t}.
\]

Since \( \ln S_{t+1} \) is normally distributed,

\[
R = \frac{\ln S_{t+1} - \tilde{\mu}}{\tilde{\sigma}}
\]

and the transformation takes the form

\(^{39}\)See Press, Teukolsky, Vetterling and Flanney[33].

\(^{40}\)This assumption is consistent with investors' expectation of the return on a stock to be positive.
$S_{t+1} = e^{R_{t+1}+\tilde{\mu}}$.  

$S_{t+1}$ is the new transformed random variable. This is the stock price at time $t+1$, generated using (5.1) to (5.6) with $S_t$ as input. In order to generate $S_{t+2}$, we again use (5.1) to (5.6) and use $S_{t+1}$ as input.

**Step 2. Generation of a time series**

We generate 190 points of a time series covering the 9 month life of an option using Step 1 above. However, the assumption of positive drift results in a upward trend in any time series generated. In order to generate a wide variety of trends, adjustment factors are introduced. A time series of 190 points is subdivided into at most 3 sections; an adjustment factor is introduced into each section of the time series: the random variable $R$ generated using the Box-Muller transformation (5.1) is augmented by the adjustment factor $F$ for that section.

$$\tilde{R} = R + F \quad (5.7)$$

where

$$F \in (-.3, -.2, -.1, 0, 1, 2, 3)$$

and the augmented variable $\tilde{R}$ is the variable that is transformed into a stock price using (5.2) to (5.6). The choice of adjustment factor to use in the generation of a particular stock price series is discussed below.

To generate a time series, the following parameters have been preset:

1) $S_0$ takes one of three preset values: 950 points, 1000 points, 1050 points. The choice is based on a preset exercise price of 1000 points.

2) The time interval $\Delta t$ is preset to 0.004 year or 1 day, assuming a 250-day year.

3) The volatility $\sigma$ takes one of 5 preset values: 0.20, 0.30, 0.40, 0.50, 0.60. One time series is generated holding the volatility constant.

4) The rate of return $r$ is preset to 0.20.
The choice of adjustment factor is based on the volatility that will be used in the generation of the stock price series. The factor $F=0$ is applied to all volatility levels. The factor $F=1$ or $F=-1$ is applied to volatility levels 0.20 and 0.30. The factor $F=2$ or $F=-2$ is applied to volatility level 0.40. The factor $F=3$ or $F=-3$ is applied to volatility levels 0.50 and 0.60. This system of adjustment ensured that there is a high probability of generating similar trends across all volatility levels given a constant set of parameters. For example, when volatility = 0.40, (5.7) becomes:

$$\tilde{R} = R + 0.2 .$$

(5.8)

The use of (5.8) creates an upward trend in a section of the stock price series.

**Step 3. Generation of the full data set for the simulation**

Holding the volatility constant, 1250 sets of option and stock time series were generated which were then screened based on a selection procedure described in Section 5.3. Selection is based on whether the time series falls within any of the specified Groups identified below.
5.3 Groups

For all time series having the same volatility, each time series is placed into one of 5 groups of options. These groups are defined based on two events: the **crossover** and the **abrupt change**.

- A **crossover** is an event such that either \( S_i \leq X \) and \( S_{i+1} > X \), or \( S_i > X \) and \( S_{i+1} \leq X \).
- An **abrupt change** is an event such that \( S_{t+1} \in \text{Abrupt Change} \).

<table>
<thead>
<tr>
<th>Group 1:</th>
<th>Crossovers</th>
<th>Abrupt Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Group 2:</td>
<td>Several</td>
<td>None</td>
</tr>
<tr>
<td>Group 3:</td>
<td>None</td>
<td>Several</td>
</tr>
<tr>
<td>Group 4:</td>
<td>Several</td>
<td>Several</td>
</tr>
<tr>
<td>Group 5:</td>
<td>Few</td>
<td>Few</td>
</tr>
</tbody>
</table>

The allocation of an option/stock time series to any of the above groups is determined by the distribution of total number of crossovers per series, the distribution of total number of abrupt changes per series and the distribution of total number of simultaneous crossovers and abrupt changes per series.

The **method of allocation of a stock price series to a group**

Given the exercise price and the Abrupt-Change-range defined in Section 4, the following numbers are known for each stock series:

\[
J_X \quad \text{total number of crossovers}
\]

\[
J_{AC} \quad \text{total number of abrupt changes}
\]

\[
J_{X&AC} \quad \text{total number of simultaneous crossovers and abrupt changes}
\]

---

\(^{41}\)See Section 4.2.2.
The distributions of $J_X$, $J_{AC}$ and $J_{X&AC}$, each based on a sample of 1250 stock series (for a given constant volatility $\sigma$), are ascertained. These distributions are used to allocate a time series of a specified volatility into one of the 5 option groups.

Given the means, $\mu_{J_X}$, $\mu_{J_{AC}}$, $\mu_{J_{X&AC}}$, and standard deviations, $\sigma_{J_X}$, $\sigma_{J_{AC}}$, $\sigma_{J_{X&AC}}$, of these distributions, a stock price series is allocated to a group if it has the properties given below for that group:

Group 1:

$$J_X < \mu_{J_X} - a\sigma_{J_X}, \quad J_{AC} < \mu_{J_{AC}} - b\sigma_{J_{AC}}, \quad \text{and} \quad J_{X&AC} < \mu_{J_{X&AC}} - c\sigma_{J_{X&AC}}$$

Group 2:

$$J_X > \mu_{J_X} + a\sigma_{J_X}, \quad J_{AC} < \mu_{J_{AC}} - b\sigma_{J_{AC}}, \quad \text{and} \quad J_{X&AC} < \mu_{J_{X&AC}} - c\sigma_{J_{X&AC}}$$

Group 3:

$$J_X < \mu_{J_X} - a\sigma_{J_X}, \quad J_{AC} > \mu_{J_{AC}} + b\sigma_{J_{AC}}, \quad \text{and} \quad J_{X&AC} < \mu_{J_{X&AC}} - c\sigma_{J_{X&AC}}$$

Group 4:

$$J_{X&AC} > \mu_{J_{X&AC}} + c\sigma_{J_{X&AC}}$$

Group 5:

$$\mu_{J_X} - a\sigma_{J_X} \leq J_X \leq \mu_{J_X} + a\sigma_{J_X} \quad \text{and} \quad \mu_{J_{AC}} + b\sigma_{J_{AC}} \leq J_{AC} \leq \mu_{J_{AC}} + b\sigma_{J_{AC}}$$

In the simulation, $a=1$, $b=1$ and $c=2$. We chose these values in order to ensure that the groups are sufficiently differentiated and that there are a large number of elements within a group. We set $c=2$ in order to ensure that all groups, except Group 4, have a low incidence of simultaneous crossovers and abrupt changes. Each group has a total of 250 options, representing 50 options for each of the 5 volatility levels. We refer to one simulation run for one option as a replication. A total of 1250 replications were done in the simulation.

In Table 5.3b, very roughly the allocation has generated actual groups of time series of the underlying stock with the following characteristics.
Table 5.3b The actual groups

<table>
<thead>
<tr>
<th></th>
<th>Crossovers</th>
<th>Abrupt Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1:</strong></td>
<td>Very few</td>
<td>Very few</td>
</tr>
<tr>
<td><strong>Group 2:</strong></td>
<td>Several</td>
<td>Very few</td>
</tr>
<tr>
<td><strong>Group 3:</strong></td>
<td>Very few</td>
<td>Several</td>
</tr>
<tr>
<td><strong>Group 4:</strong></td>
<td>Several</td>
<td>Several</td>
</tr>
<tr>
<td><strong>Group 5:</strong></td>
<td>Few</td>
<td>Few</td>
</tr>
</tbody>
</table>

5.4 From set-up to wind-down

In this section, we discuss the mechanics of hedging, from setting up to winding down the hedge. We consider delta hedging and 7 minimax hedging strategies. These strategies involve rebalancing the hedge at uniform intervals of time; in the simulation, the interval is one day. Daily data include the stock price, the option price, the risk-free interest rate, the time to maturity and the volatility. The risk-free interest rate is preset to 0.10. Dividends are excluded from the analysis, and $N$, the number of contracted shares, is 100.

5.4.1 Setting up the hedge

At time 0, under each strategy the hedger holds the same number of shares ($n_0$) based on delta, and the same initial cumulative value of cash inflow minus cash outflow given by Eqn (4.25).

5.4.2 Rebalancing the hedge

Every day through to the maturity date $T$, the hedge is rebalanced on the basis of the strategies' recommendations. The trajectory of the number of shares held at time $t$, $n_t$, varies with the hedging strategy used. The actual cumulative value of cash inflow minus cash outflow at time $t$ is given by Eqn (4.24).

---

42This is the continuous rate.
5.4.3 Winding down the hedge

At time $T$, if the holder does not exercise his option to buy the shares, each strategy disposes of its portfolio in the same way: selling any shares held, or buying any shares sold short, at time $T-1$ at $S_T$.

5.5 Stratification

The simulation has the following stratification based on volatility (sigma), group, and minimax variant.

- Levels of Sigma: 0.2, 0.3, 0.4, 0.5, 0.6
- Groups: 1, 2, 3, 4, 5
- Minimax variants: A, B, C, D, D*, E, F
5.6 Results of the simulation study

We give the results of this simulation mainly in terms of the performance and the relative performance of the minimax strategy. We define the performance of a strategy as the final cumulative value of cash inflow minus cash outflow in using that strategy on an option, standardized as a percentage of the notional contract value of that option. We define relative performance of a minimax variant as the performance of that variant minus the performance of delta hedging (DH). In the following commentary on this simulation, we say that a variant outperforms DH if its performance is higher than that of DH and, for any group of options, the difference for that group is significant at the 1% level.

In Table 5.6.1 we give the relative performance of the seven minimax variants, and below each difference we list, in italics in small font, the absolute value of the t-statistics, followed by ** if the difference is significant at the 2% level, and * if it is significant at the 10% level. In this table the main distributor is the group, as defined in Section 5.3, and the sub-distributor is sigma, as defined in Section 5.5. We give two other tables, both derived from Table 5.6.1. Table 5.6.2 gives the difference in relative performance between selected groups, with sigma as the main distributor, and variant as the sub-distributor. Table 5.6.3 gives the difference in relative performance between selected variants, with group as the main distributor, and sigma as the sub-distributor.

---

43 A strategy may be either a minimax variant or delta hedging.
44 This is calculated after winding down the hedge.
45 Following Samuelson, all cells in Table 5.6.1 have been standardized by dividing the original profit by the exercise price. In the simulation study, we set the exercise price at X=1000; this makes the original profit effectively standardized. In the empirical study, because of the differences in exercise prices, the standardization becomes relevant.
46 The notional value of the contract is the number of contracted shares multiplied by the exercise price. The total is the summation over all notional values.
47 We use significance levels of 2% and 10% to express the results of a 2-tailed test; however, we use a 1-tailed test for the sign of the difference at the 1% level.
The tables give the results for all groups, including Group 5 as an illustration only. We do not include Group 5 in the analysis because it does not represent any well-defined characteristics of the time series of stock prices.

5.6.1 Performance of minimax variants against delta hedging

In Table 5.6.1, we give the relative performance of the seven variants for each group, and within each group, for each level of sigma. We comment on the main results in Section 5.6.1.1.
Table 5.6.1 Relative performance of variants by group and by sigma. (*-values in italics)

Units: Percentage points of final cumulative value of cash inflow minus cash outflow divided by the notional value of the contract

<table>
<thead>
<tr>
<th>Group</th>
<th>Sigma</th>
<th>A-DH</th>
<th>B-DH</th>
<th>C-DH</th>
<th>D-DH</th>
<th>D*-DH</th>
<th>E-DH</th>
<th>F-DH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.7</td>
<td>0.2</td>
<td>-0.6</td>
<td>5.1</td>
<td>7.5</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.4</td>
<td>-0.1</td>
<td>0.2</td>
<td>-0.3</td>
<td>5.0</td>
<td>6.1</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.2</td>
<td>-1.0</td>
<td>0.1</td>
<td>-1.0</td>
<td>4.9</td>
<td>6.7</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1</td>
<td>-1.6</td>
<td>0.1</td>
<td>-1.5</td>
<td>4.3</td>
<td>5.7</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.1</td>
<td>-1.7</td>
<td>0.1</td>
<td>-1.6</td>
<td>4.2</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>0.6</td>
<td>-0.7</td>
<td>1.3</td>
<td>-2.1</td>
<td>5.8</td>
<td>9.1</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.4</td>
<td>-0.7</td>
<td>0.2</td>
<td>-2.0</td>
<td>4.6</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.5</td>
<td>-2.0</td>
<td>0.1</td>
<td>-1.6</td>
<td>6.4</td>
<td>8.0</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.2</td>
<td>-0.1</td>
<td>0.0</td>
<td>-0.7</td>
<td>5.8</td>
<td>6.8</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.3</td>
<td>-0.1</td>
<td>0.1</td>
<td>-1.5</td>
<td>5.8</td>
<td>5.7</td>
<td>6.4</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.1</td>
<td>-0.1</td>
<td>1.4</td>
<td>1.2</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>-0.2</td>
<td>-0.7</td>
<td>0.2</td>
<td>-1.5</td>
<td>1.8</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>-0.2</td>
<td>-2.4</td>
<td>0.0</td>
<td>-2.2</td>
<td>0.5</td>
<td>3.2</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-0.3</td>
<td>-0.0</td>
<td>0.0</td>
<td>-2.8</td>
<td>-0.1</td>
<td>3.4</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>-0.4</td>
<td>-3.9</td>
<td>0.0</td>
<td>-3.2</td>
<td>0.1</td>
<td>3.8</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>0.7</td>
<td>-0.7</td>
<td>1.6</td>
<td>1.9</td>
<td>7.9</td>
<td>10.4</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.6</td>
<td>-0.7</td>
<td>0.3</td>
<td>1.4</td>
<td>6.4</td>
<td>7.2</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>-0.7</td>
<td>-0.7</td>
<td>0.3</td>
<td>3.4</td>
<td>9.3</td>
<td>9.3</td>
<td>11.3</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.8</td>
<td>-2.0</td>
<td>0.1</td>
<td>4.9</td>
<td>10.5</td>
<td>8.8</td>
<td>11.9</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.8</td>
<td>-2.0</td>
<td>0.3</td>
<td>5.3</td>
<td>11.1</td>
<td>7.7</td>
<td>11.8</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>0.3</td>
<td>-0.3</td>
<td>0.1</td>
<td>-0.3</td>
<td>4.4</td>
<td>6.0</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.3</td>
<td>-0.7</td>
<td>0.0</td>
<td>-0.6</td>
<td>2.6</td>
<td>4.5</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>-0.2</td>
<td>-0.2</td>
<td>0.0</td>
<td>0.3</td>
<td>3.2</td>
<td>5.6</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-0.2</td>
<td>-2.1</td>
<td>0.0</td>
<td>-1.8</td>
<td>2.0</td>
<td>5.6</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>-0.2</td>
<td>-1.9</td>
<td>0.1</td>
<td>-1.3</td>
<td>4.1</td>
<td>4.7</td>
<td>4.0</td>
</tr>
</tbody>
</table>
5.6.1.1 Main results from Table 5.6.1

In this section we give the main results from Table 5.6.1.

A-DH, C-DH: The relative performances of variants C and A are highly correlated; both perform much the same as DH\(^48\). C is not significantly different from DH for any group or for any level of sigma; A outperforms DH only for Group 4.

B-DH, D-DH: The relative performances of B and D are very highly correlated, even more highly than A and C. For all groups except Group 4 and for all levels of sigma, DH outperforms B and D by roughly equal amounts; for Group 4 for all levels of sigma, B and D outperform DH, again by roughly equal amounts.

D*-DH: For all groups, except Group 3, and for all levels of sigma, D* outperforms DH; the rank order of the relative performance is Group 4, Group 1 and Group 2.

E-DH: For all groups and all levels of sigma, E outperforms DH; the rank order of the relative performance is Group 4, Group 2, Group 1 and Group 3.

F-DH: For all groups and all levels of sigma, F outperforms DH; as for E-DH, the rank order of the relative performance is Group 4, Group 2, Group 1 and Group 3.

---

\(^48\)We tested the hypothesis that the number of shares recommended by the 95%-Level variants will be significantly different from the number of shares recommended by delta hedging. Because delta hedging ensures that any gain or loss in the stock is offset by any loss or gain in the option for a small change in the stock price for a small interval of time and because for the 95%-Level variants of minimax we minimized the hedging error within the 95% confidence interval, and as we move the bounds closer we increasingly approximate delta hedging, we anticipate that the hypothesis will be rejected for the 95%-Level variants. The correlation coefficients between the 95%-Level variants and DH vary from 0.93 to 0.99.
5.6.2 Relative performance between groups

Table 5.6.2 is a reworking of Table 5.6.1; it gives the differences between the relative performances of selected variants. The only new information it provides is the t-statistic for each difference. Because the performances of A and C differ so little from DH, and because the performances of B and D are very highly correlated, we confine the subsequent discussion to D, D*, E and F, noting that results for D apply equally to B.

We use the following abbreviations: GI (for Group 1), G2 (for Group 2), G3 (for Group 3) and G4 (for Group 4).

Table 5.6.2: Difference in relative performance between groups. *(t-values in italics)*

<table>
<thead>
<tr>
<th>Sigma</th>
<th>Variant</th>
<th>G4-G1</th>
<th>G4-G2</th>
<th>G4-G3</th>
<th>G2-G3</th>
<th>G2-G1</th>
<th>G3-G1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>D</td>
<td>2.5</td>
<td>4.0</td>
<td>3.0</td>
<td>-1.5</td>
<td>7.3**</td>
<td>5.1**</td>
</tr>
<tr>
<td></td>
<td>D*</td>
<td>2.8</td>
<td>2.1</td>
<td>6.5</td>
<td>0.7</td>
<td>4.8</td>
<td>3.8**</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>2.9</td>
<td>1.3</td>
<td>9.2</td>
<td>1.6</td>
<td>7.9</td>
<td>6.4**</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>3.2</td>
<td>1.5</td>
<td>9.5</td>
<td>1.7</td>
<td>8.0</td>
<td>6.3**</td>
</tr>
<tr>
<td>0.3</td>
<td>D</td>
<td>1.7</td>
<td>3.4</td>
<td>2.9</td>
<td>-1.7</td>
<td>8.2**</td>
<td>6.3**</td>
</tr>
<tr>
<td></td>
<td>D*</td>
<td>1.4</td>
<td>1.8</td>
<td>4.0</td>
<td>0.1</td>
<td>3.0</td>
<td>3.2**</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1.1</td>
<td>1.0</td>
<td>6.2**</td>
<td>0.9</td>
<td>6.7**</td>
<td>6.3**</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.6</td>
<td>1.8</td>
<td>4.8</td>
<td>-0.2</td>
<td>3.0</td>
<td>3.1**</td>
</tr>
<tr>
<td>0.4</td>
<td>D</td>
<td>4.4</td>
<td>5.0</td>
<td>5.6</td>
<td>-0.6</td>
<td>8.6**</td>
<td>6.3**</td>
</tr>
<tr>
<td></td>
<td>D*</td>
<td>4.4</td>
<td>2.9</td>
<td>8.8</td>
<td>1.5</td>
<td>5.9</td>
<td>3.9**</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>2.6</td>
<td>1.3</td>
<td>6.1</td>
<td>1.3</td>
<td>4.8</td>
<td>6.4**</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>4.8</td>
<td>2.9</td>
<td>9.5</td>
<td>1.9</td>
<td>6.6</td>
<td>6.6**</td>
</tr>
<tr>
<td>0.5</td>
<td>D</td>
<td>6.4</td>
<td>6.5</td>
<td>7.7</td>
<td>-0.1</td>
<td>1.3</td>
<td>1.3**</td>
</tr>
<tr>
<td></td>
<td>D*</td>
<td>6.2</td>
<td>4.7</td>
<td>10.6</td>
<td>1.5</td>
<td>5.9</td>
<td>4.4**</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>3.1</td>
<td>2.0</td>
<td>5.4</td>
<td>1.1</td>
<td>3.4</td>
<td>2.3**</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>6.5</td>
<td>4.8</td>
<td>10.7</td>
<td>1.7</td>
<td>5.9</td>
<td>4.2**</td>
</tr>
<tr>
<td>0.6</td>
<td>D</td>
<td>6.9</td>
<td>6.8</td>
<td>8.5</td>
<td>0.0</td>
<td>1.7</td>
<td>1.6**</td>
</tr>
<tr>
<td></td>
<td>D*</td>
<td>6.9</td>
<td>5.3</td>
<td>11.0</td>
<td>1.6</td>
<td>5.7</td>
<td>4.1**</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>3.0</td>
<td>2.0</td>
<td>3.9</td>
<td>1.0</td>
<td>1.9</td>
<td>0.9**</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>7.1</td>
<td>5.4</td>
<td>10.3</td>
<td>1.7</td>
<td>4.8</td>
<td>3.1**</td>
</tr>
</tbody>
</table>
5.6.2.1 Main results from Table 5.6.2

In this section, we give the main results from Table 5.6.2. We compare the performance of each selected variant when it is applied to options in two different groups. We say that one group outperforms another group if the relative performance of a variant applied to that group is higher than its relative performance when it is applied to the other group, and the difference in relative performance is significant at the 1% level.

G2-G1: For most levels of sigma, G2 outperforms G1 for F, D* and E (given in rank order); for D, G1 outperforms G2 for low levels of sigma.

G2-G3: For most levels of sigma, G2 outperforms G3 for F, D* and E (given in rank order); for D, G3 outperforms G2 for lower levels of sigma.

G3-G1: For all levels of sigma, G1 outperforms G3 for F, D*, D and E (given in rank order).

G4-G1: For all sigma, G4 outperforms G1 for F, D, D*, and E (given in rank order).

G4-G2: As for G4-G1, for all levels of sigma, G4 outperforms G2. However in contrast to G4-G1, D has the highest rank, the rank order of the differences being D, F, D* and E.

G4-G3: As for G4-G1, for all levels of sigma, G4 outperforms G3, with the rank order of the differences being F, D*, E and D.
5.6.3 Relative performance between variants

In this section, we contrast the relative performance of selected variants applied to different groups of options at different levels of sigma. Table 5.6.3 is a reworking of Table 5.6.1, with variants A and B excluded. C is included to show its contrast with its counterpart variant E, which includes transaction costs in its objective function.

Table 5.6.3 Difference in relative performance between variants. (t-values in italics)

Units: Percentage points of final cumulative value of cash inflow minus cash outflow divided by the notional value of the contract.

<table>
<thead>
<tr>
<th>Group</th>
<th>Sigma</th>
<th>F-E</th>
<th>E-C</th>
<th>F-D</th>
<th>D*-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2</td>
<td>1.4</td>
<td>7.3</td>
<td>8.5</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.3</td>
<td>5.9</td>
<td>6.7</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>-0.2</td>
<td>6.6</td>
<td>6.6</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-0.3</td>
<td>5.7</td>
<td>6.9</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.0</td>
<td>4.7</td>
<td>6.3</td>
<td>5.8</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>0.5</td>
<td>8.8</td>
<td>11.7</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.4</td>
<td>7.8</td>
<td>10.0</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.3</td>
<td>6.7</td>
<td>8.6</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.7</td>
<td>5.7</td>
<td>7.9</td>
<td>7.3</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0.4</td>
<td>1.2</td>
<td>2.7</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.3</td>
<td>3.2</td>
<td>4.7</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>-1.4</td>
<td>3.2</td>
<td>4.1</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-2.2</td>
<td>3.4</td>
<td>4.0</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>-2.3</td>
<td>3.8</td>
<td>4.7</td>
<td>3.3</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>0.7</td>
<td>10.1</td>
<td>9.2</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.8</td>
<td>6.9</td>
<td>6.6</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>2.0</td>
<td>9.0</td>
<td>7.9</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>3.1</td>
<td>8.5</td>
<td>7.0</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>4.1</td>
<td>7.4</td>
<td>6.5</td>
<td>5.8</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>0.3</td>
<td>5.8</td>
<td>6.6</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>-0.1</td>
<td>4.5</td>
<td>5.0</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>-0.6</td>
<td>5.5</td>
<td>6.3</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-1.7</td>
<td>5.6</td>
<td>5.7</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>-1.7</td>
<td>4.6</td>
<td>2.7</td>
<td>5.4</td>
</tr>
</tbody>
</table>

143
5.6.3.1 Main results from Table 5.6.3
In this section we give the main results from Table 5.6.3.

E-C: E, the counterpart variant of C, includes transaction costs in its objective function. For all groups and all levels of sigma, E outperforms C, with the rank order of the differences being G4, G2, G1 and G3.

F-D: F, the counterpart variant of D, includes transaction costs in its objective function. For all groups and all levels of sigma, F outperforms D. Compared to E-C, the rank order position of G4 and G2 is reversed, with the rank order of the differences being G2, G4, G1 and G3.

F-E: E and F have the same objective function; F is the Abrupt-Change variant and E is the 95%-Level variant. For all levels of sigma, F outperforms E for G4, and E outperforms F for G3. There is no systematic pattern in the performance for G1 and G2, and any differences are small.

D*-D: D* is the weighted variant of D. For all groups and all levels of sigma, D* outperforms D, with the rank order of the differences being G2, G1, G4 and G3.

5.7 Statistical testing of hypotheses

In this section we formulate and test hypotheses; we also comment on any rejected hypotheses. Using the differences and t-values in Tables 5.6.1 and 5.6.2, we test sets of hypotheses for variants D, D*, E and F. We test the significance of a positive or a negative difference at the 1% level. For the reason given in Section 5.6, we do not consider A, B and C. In the following, our hypotheses are formulated for each level of sigma. However, for compactness we summarize the results of our tests on the basis that

49For statistical hypothesis testing, see Bury[4].
a composite hypothesis, over all levels of sigma, will be accepted if at least 4 of the 5 hypotheses, one for each of the 5 sigma levels, are accepted.

5.7.1 Hypotheses on the relative performance of the variants

Because minimax is designed to perform better when the underlying stock has crossovers with respect to the exercise price and has abrupt changes, we hypothesize (H2, H3 and H4) that D, D*, E and F will outperform DH for those groups characterized by crossovers or abrupt changes (Groups 4, 3 and 2) and we hypothesize (H1) that DH will outperform these variants for Group 1. Using the differences and t-values in Table 5.6.1, we tested the following hypotheses for each variant, and give the result Accept only if at least 4 of the 5 hypotheses for that variant applied to the five levels of sigma are accepted.

Table 5.7.1 Composite results of tests on the relative performance of the variants.

(R=Reject)

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>D</th>
<th>D*</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 DH will outperform the variant for Group 1</td>
<td>Accept</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>H2 the variant will outperform DH for Group 2</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H3 the variant will outperform DH for Group 3</td>
<td>R</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H4 the variant will outperform DH for Group 4</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
</tbody>
</table>

5.7.1.1 Comments on rejected hypotheses

For H1, because D*, E and F constrain transaction costs and have lower variations in their hedge recommendations, the consequential savings in transaction costs may explain the rejection of H1.

For H3, because the hedge recommendations of D are based on possible abrupt changes, they are highly volatile; further they are not constrained for transaction costs. As a result, D may have very high transaction costs that outweigh any cost saving that D may achieve in anticipating abrupt changes.
Hypotheses on relative performance between groups

Because minimax is designed to perform best when the underlying stock has crossovers and, to a lesser extent, also has abrupt changes, we hypothesize (H8, H9 and H10) that for each variant and for each sigma the options in Group 4, which include several crossovers and abrupt changes, will outperform the options in the other groups. Because Group 2 contains several crossovers, we hypothesize (H5) that for each variant and for each sigma the options in Group 2 will outperform the options in Group 1. Because Group 3 contains several abrupt changes, we hypothesize (H7) that for each variant and for each sigma the options in Group 3 will outperform the options in Group 1.

Because Group 2 contains several crossovers and very few abrupt changes, and Group 3 contains very few crossovers and several abrupt changes, and because a crossover implies a significant change in the value of an option whereas an abrupt change does not necessarily imply a significant change (for example, deep-out-of-the-money options), and finally because hedging errors are generally higher for at-the-money options than they are for deep-in-the-money or deep-out-of-the-money options, we hypothesize (H6) that for each variant and for each sigma the options in Group 2 will outperform the options in Group 3.

Using the differences and t-values in Table 5.6.2, we tested the following hypotheses. For compactness, we give the results of the six composite hypotheses, accepting the composite if at least 4 out of the 5 hypotheses at different levels of sigma are accepted.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>D</th>
<th>D*</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>H5</td>
<td>R</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H6</td>
<td>R</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H7</td>
<td>R</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H8</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H9</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H10</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
</tbody>
</table>
5.7.2.1 Comments on rejected hypotheses

For H5 and H6, for variant D, because the hedge recommendations of D are highly volatile, and are not constrained for transaction costs, D may have very high transaction costs that outweigh any cost saving that D may achieve in anticipating abrupt changes, even if G2 has crossovers.

For H7, for all variants, because G3 consists mainly of options that are either deep-in-the-money or deep-out-of-the-money, the anticipation by D, D*, E and F of a worst case, even if it does occur, does not result in a saving because there is no significant change in the state of the option that would justify the hedge recommendations of the variants.

5.7.3 Hypotheses on the relative performance between variants

Because E and F, the variants with transaction cost terms in the objective function, constrain transaction costs, and C and D do not, we hypothesize (H11 and H12) that E and F will outperform their counterparts C and D. Because F, the Abrupt-Change variant, is designed to anticipate abrupt changes that may lead to crossovers, we hypothesize (H13a) that F will outperform E for G2, G3 and G4. Because G1 has few crossovers or abrupt changes, we hypothesize (H13b) that there will be no significant difference between F and E for G1. Because D*, the weighted variant of D, is designed to constrain transaction costs, we hypothesize (H14) that D* will outperform D.

Using the differences and t-values in Table 5.6.3, we tested the following hypotheses. For the reason given in Section 5.6.2, we do not consider A, B and C. For compactness (see 5.7.1), we give the results of the four composite hypotheses applied to 4 groups.
Table 5.7.3 Composite results of tests on the relative performance between variants. (R=Reject, n.a.=not applicable)

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
</tr>
</thead>
<tbody>
<tr>
<td>H11 E will outperform C</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H12 F will outperform D</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H13a F will outperform E for G2, G3, G4</td>
<td>n.a.</td>
<td>R</td>
<td>R</td>
<td>Accept</td>
</tr>
<tr>
<td>H13b F will not outperform E and vice versa</td>
<td>Accept</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>H14 D* will outperform D</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
</tbody>
</table>

5.7.3.1 Comments on rejected hypotheses

Because the 95%-Level price range in E often includes the exercise price, the computation of the objective function often includes options that are in all states (in-the-money, at-the-money, and out-of-the-money). Given this, and because G2 has several crossovers, the states of the options implicitly anticipated under E may be similar to the states explicitly anticipated under F. The rejection of the hypothesis H13a for G2 suggests the states were similar. For G3, we give the same reason as for the rejection of H7 (Section 5.7.2).

5.8 Intersigma performance of D*, E and F

Table 5.8 is a minimal reworking of Table 5.6.1 for variants D*, E and F, with sigma now the main distributor, and groups the sub-distributor. These three variants are the only ones that systematically outperform DH, and we refer to them as the high performing variants.
Table 5.8 Differences in relative performance between high performing variants.  
(t-values in italics)

Units: Percentage points of final cumulative value of cash inflow minus cash outflow divided by the notional value of the contract.

<table>
<thead>
<tr>
<th>Sigma</th>
<th>Group</th>
<th>D*-DH</th>
<th>E-DH</th>
<th>F-DH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1</td>
<td>5.1</td>
<td>7.5</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5.8</td>
<td>9.1</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.4</td>
<td>1.2</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>7.9</td>
<td>10.4</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4.4</td>
<td>6.0</td>
<td>6.3</td>
</tr>
<tr>
<td>0.3</td>
<td>1</td>
<td>5.0</td>
<td>10.6</td>
<td>15.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.6</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.8</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6.4</td>
<td>7.2</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.6</td>
<td>4.5</td>
<td>4.4</td>
</tr>
<tr>
<td>0.4</td>
<td>1</td>
<td>4.9</td>
<td>6.7</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.4</td>
<td>8.0</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.5</td>
<td>3.2</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>9.3</td>
<td>9.3</td>
<td>11.3</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.2</td>
<td>5.6</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>4.3</td>
<td>5.7</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5.8</td>
<td>6.8</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.1</td>
<td>3.4</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10.5</td>
<td>8.8</td>
<td>11.9</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.0</td>
<td>5.6</td>
<td>3.9</td>
</tr>
<tr>
<td>0.6</td>
<td>1</td>
<td>4.2</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5.8</td>
<td>5.7</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1</td>
<td>3.8</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>11.1</td>
<td>7.7</td>
<td>11.8</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4.1</td>
<td>4.7</td>
<td>4.0</td>
</tr>
</tbody>
</table>

149
5.8.1 Main results from the analysis of the intersigma performance of D*, E and F

Figure 5.8.1a gives a graphical representation of the performance of the high performing variants averaged over all levels of sigma. The x-axis gives the average distance of a particular group from the exercise price. Each group consists of 50 options or stock price series, corresponding to 50 replications; each stock price series consists of 190 daily prices. The average squared deviation from the exercise price over 9500 (190*50) prices was calculated; the square root gives the average distance from the exercise price for that group.

\[
\text{average distance from the exercise price} = \sqrt{\frac{\sum_{i=1}^{50} \sum_{l=1}^{190} (S_l^i - X)^2}{190*50}} \tag{5.9}
\]

where \(i\) refers to a particular time series in a group and \(l\) refers to a particular price. The average distance from the exercise price represents the degree of moneyness\(^{50}\) of the options in that group. Because, in general, sigma level 0.4 is representative of the other sigma levels, we illustrate in Figure 5.8.1a the effect of sigma and moneyness, represented by the distance from the exercise price, using the performance for sigma level 0.4 only.

In Figure 5.8.1a, there is a consistent relative ordering of groups along the x-axis. Groups 4 and 2 are near the exercise price; accordingly, these two groups consist of at-the-money options. Group 3 is farthest from the exercise price; this indicates that it consists of either deep-in-the-money or deep-out-of-the-money options. The relative ordering of either Group 2 or 4 and of Group 3 was by construction: Groups 2 and 4 consists of time series with many crossovers, and therefore close to the exercise price, and Group 3 consists of time series with no crossovers, and therefore away from the exercise price. As noted earlier, D*, E and F consistently outperform delta hedging for each group.

\(^{50}\)The amount that the option is in-the-money or out-of-the-money.
In Figure 5.8.1b, we show the variation with sigma in the performance of the high performing variants E, F and D*. E uses the 95%-Level range to define the worst case scenario, F uses the Abrupt-Change range, and D* also uses the Abrupt-Change range together with a heuristic weighting system to give weighted hedge recommendations. For each of five levels of sigma, a regression line shows the relationship between performance and degree of moneyness, represented by the average distance from the exercise price. All five regressions for each of the variants have a negative slope.

Figure 5.8.1b shows that the higher the volatility, the wider the range of distances over which the variants outperform DH. For sigma levels 0.2 and 0.3, the three variants
significantly outperform DH for distances close to the exercise price, but for distances further from the exercise price, DH outperforms the variants. In contrast, for sigma levels 0.5 and 0.6, the variants significantly outperform DH for a wide range of distances from the exercise price.

We note the regression lines for D* have the fewest intersections, and for E they have the most intersections. We interpret this difference in the degree of parallelism below.

For D*, the lines have few (3) intersections; this degree of parallelism is consistent with D* outperforming DH more for higher values of sigma than for lower values of sigma at all average distances from the exercise price. For E, the lines have many (7) intersections; this degree of parallelism is consistent with the result that when the average distance is small, E outperforms DH more for lower values of sigma than it does for higher values of sigma, and when the average distance is large, that relationship between performance and sigma is reversed. For high values of sigma, E outperforms DH by roughly the same amount for a wide range of moneyness. For F, the lines have 6 intersections, and F is intermediate between D* and E.

Compared to E and F, D* outperforms DH over a small range of distance. This implies that D* performs well only for a limited range of moneyness. Because F outperforms E and D* when the average distance is small, F is the most suitable variant for at-the-money options. Because E outperforms F when the average distance is large, F is not suitable for deep-out-of-the-money or deep-in-the-money options; E is more suitable for such options.
Figure 5.8b Variation with sigma of the high performing variants, E, F and D*.
6 Empirical study of the performance of different strategies

In this section, we present a limited empirical study of hedging the risk of writing a European call option\textsuperscript{51} for 30 options available in the UK options market. The writer of the call incurs a potential liability in case of exercise of the option by the buyer, and receives a premium; he is obliged to offer the buyer $N$ shares of the stock at the exercise price, $X$, at the execution date. In the illustrations, $N=1000$. In Section 6.2, we describe the mechanics of the hedging from set-up to wind-down. In Section 6.3, we give the main results of the study, and in Section 6.4, we present and test hypotheses and compare their states to that of similar hypotheses tested in the simulation study.

6.1 The objective of the empirical study

This limited empirical study is designed to illustrate the performance of the minimax variants when they are applied to real data, and to ascertain the extent to which the results of the simulation study, and the state of the hypotheses tested during it, are supported by the analysis of real data.

6.2 From set-up to wind-down

In Section 5.4, we discussed the mechanics of hedging, from setting up to winding down the hedge. For this empirical study, we use the same mechanics with a few changes in parameters. The strategies involve rebalancing the hedge at uniform intervals of time; in the illustration, the interval is one week, i.e. 5 trading days. Weekly data\textsuperscript{52} include the market price of the stock, the market price of the option, the risk-free interest rate, the time to maturity and an estimate of the volatility of the return on the stock. The estimates of volatility used in the illustration are based on the most recent 100 days of stock price movement and on implied volatility\textsuperscript{53}. The risk-free interest rate is based on the discount

\textsuperscript{51}Since $k=1$, the superscript $i$ will be dropped.

\textsuperscript{52}Data were supplied by Datastream, International. All prices are mid-prices. The stock price series used were Datastream's Adjusted Stock Price Series: these are the original stock price series with dividend adjustments.

\textsuperscript{53}Implied volatility is the volatility implied by an option price observed in the market.
rate\textsuperscript{54} provided by a Treasury Bill that expires at about the same time as the option. In the illustration, \( N \), the number of contracted shares, is 1000.

6.3 Results of the empirical study

The strategies defined in Section 4.7 were used to hedge the risk of writing one call option; 30 options were used in the study. The universe of calls from which the calls were selected is all call options written in the London market from 1990 to 1992 that had a life of at least 30 weeks; the calls were not strictly randomly selected, being selected on the basis of the ready availability of data for them. For the 30 options selected, 7 were allocated to a group that had several crossovers\textsuperscript{55} and abrupt changes\textsuperscript{56}, while 6 were allocated to a group that had very few crossovers or abrupt changes. In our analysis we concentrate on those two groups, and compare them to Group 4 and Group 1 of the simulation study. However, we also give some results for the 30 options.

6.3.1 Relative performance of minimax variants, with intergroup comparisons

We present in Table 6.3.1a the relative performance for 30 options. We then form two groups of options: Group 1E (abbreviated to G1E) and Group 4E (abbreviated to G4E) that very roughly correspond to Group 1 and Group 4 in the simulation study. The stock price series in Group 1E have very few crossovers or abrupt changes, and in Group 4E they have several crossovers and abrupt changes. The number of options in G1E and G4E is small, respectively, 6 and 7.

\textsuperscript{54}The discount rate was used to find the current value of the Treasury Bill based on a face value of £1 and the time to maturity. The risk-free interest rate is the continuous rate that discounts the face value to the computed current value.

\textsuperscript{55}A crossover is a switch in the state of the option. See Section 4.2.2.

\textsuperscript{56}An abrupt change is a fairly large movement in stock price. See Section 4.2.2.
Table 6.3.1a Relative performance of the variants for 30 options.
Units: Percentage points of final cumulative value of cash inflow minus cash outflow divided by the notional value of the contract

<table>
<thead>
<tr>
<th>Variant</th>
<th>A-DH</th>
<th>B-DH</th>
<th>C-DH</th>
<th>D-DH</th>
<th>D*-DH</th>
<th>E-DH</th>
<th>F-DH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tesco</td>
<td>-2.5</td>
<td>-3.6</td>
<td>-0.4</td>
<td>-11.6</td>
<td>-5.7</td>
<td>1.4</td>
<td>-2.6</td>
</tr>
<tr>
<td>Boots</td>
<td>-0.7</td>
<td>0.5</td>
<td>-0.1</td>
<td>-6.7</td>
<td>-0.7</td>
<td>1.7</td>
<td>1.5</td>
</tr>
<tr>
<td>Sainsbury</td>
<td>-0.6</td>
<td>-2.3</td>
<td>-0.1</td>
<td>-13.1</td>
<td>-3.3</td>
<td>2.4</td>
<td>-3.2</td>
</tr>
<tr>
<td>BAir</td>
<td>-1.1</td>
<td>-2.9</td>
<td>-0.1</td>
<td>-3.9</td>
<td>-0.8</td>
<td>1.5</td>
<td>2.6</td>
</tr>
<tr>
<td>GEC</td>
<td>-0.8</td>
<td>-4.2</td>
<td>-0.2</td>
<td>-14.0</td>
<td>-9.2</td>
<td>0.2</td>
<td>-4.8</td>
</tr>
<tr>
<td>Allied</td>
<td>-0.1</td>
<td>-1.7</td>
<td>-0.1</td>
<td>-8.1</td>
<td>1.1</td>
<td>3.5</td>
<td>3.7</td>
</tr>
<tr>
<td>Ladbrok 1</td>
<td>0.5</td>
<td>-1.3</td>
<td>-0.1</td>
<td>-6.5</td>
<td>4.2</td>
<td>-2.4</td>
<td>7.2</td>
</tr>
<tr>
<td>Ladbrok 2</td>
<td>0.1</td>
<td>-2.9</td>
<td>-0.2</td>
<td>-8.0</td>
<td>-6.3</td>
<td>-2.3</td>
<td>-4.9</td>
</tr>
<tr>
<td>Cadbury 1</td>
<td>-1.0</td>
<td>-4.8</td>
<td>-0.3</td>
<td>-5.6</td>
<td>0.2</td>
<td>2.2</td>
<td>2.5</td>
</tr>
<tr>
<td>Cadbury 2</td>
<td>2.0</td>
<td>1.8</td>
<td>0.8</td>
<td>-1.7</td>
<td>6.0</td>
<td>7.8</td>
<td>9.3</td>
</tr>
<tr>
<td>Hanson 1</td>
<td>2.0</td>
<td>4.3</td>
<td>0.6</td>
<td>1.0</td>
<td>9.0</td>
<td>2.5</td>
<td>9.9</td>
</tr>
<tr>
<td>Hanson 2</td>
<td>-1.9</td>
<td>-3.1</td>
<td>-0.2</td>
<td>-6.2</td>
<td>-4.3</td>
<td>1.2</td>
<td>-0.4</td>
</tr>
<tr>
<td>P &amp; O</td>
<td>1.6</td>
<td>2.8</td>
<td>0.7</td>
<td>0.0</td>
<td>6.6</td>
<td>4.5</td>
<td>10.3</td>
</tr>
<tr>
<td>Vodafone 1</td>
<td>0.2</td>
<td>-2.7</td>
<td>-0.1</td>
<td>-10.1</td>
<td>-5.7</td>
<td>1.5</td>
<td>-3.3</td>
</tr>
<tr>
<td>Vodafone 2</td>
<td>0.2</td>
<td>1.7</td>
<td>0.0</td>
<td>-1.9</td>
<td>2.8</td>
<td>1.8</td>
<td>5.8</td>
</tr>
<tr>
<td>Prudential 1</td>
<td>4.3</td>
<td>10.3</td>
<td>1.5</td>
<td>2.6</td>
<td>13.1</td>
<td>7.6</td>
<td>14.2</td>
</tr>
<tr>
<td>Prudential 2</td>
<td>1.2</td>
<td>3.3</td>
<td>0.6</td>
<td>-4.1</td>
<td>9.7</td>
<td>5.3</td>
<td>11.3</td>
</tr>
<tr>
<td>Marks &amp; Spencer</td>
<td>0.9</td>
<td>5.0</td>
<td>0.4</td>
<td>3.1</td>
<td>8.9</td>
<td>1.7</td>
<td>11.9</td>
</tr>
<tr>
<td>Shell 1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2</td>
<td>2.8</td>
<td>6.9</td>
<td>2.3</td>
<td>7.6</td>
</tr>
<tr>
<td>Shell 2</td>
<td>0.1</td>
<td>-0.7</td>
<td>0.1</td>
<td>-11.5</td>
<td>-3.0</td>
<td>3.3</td>
<td>1.6</td>
</tr>
<tr>
<td>Eurotunnel</td>
<td>2.6</td>
<td>3.2</td>
<td>0.7</td>
<td>1.0</td>
<td>4.8</td>
<td>4.4</td>
<td>6.2</td>
</tr>
<tr>
<td>Glaxo</td>
<td>1.7</td>
<td>4.2</td>
<td>0.6</td>
<td>2.7</td>
<td>8.4</td>
<td>2.8</td>
<td>7.2</td>
</tr>
<tr>
<td>Guinness 1</td>
<td>1.5</td>
<td>-0.6</td>
<td>0.7</td>
<td>-9.8</td>
<td>0.6</td>
<td>4.2</td>
<td>5.5</td>
</tr>
<tr>
<td>Guinness 2</td>
<td>3.1</td>
<td>6.1</td>
<td>0.9</td>
<td>6.2</td>
<td>11.3</td>
<td>5.2</td>
<td>10.7</td>
</tr>
<tr>
<td>Forte</td>
<td>0.3</td>
<td>-2.8</td>
<td>-0.2</td>
<td>-0.4</td>
<td>1.4</td>
<td>-2.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Thames Water</td>
<td>-0.6</td>
<td>-2.3</td>
<td>-0.1</td>
<td>3.0</td>
<td>1.0</td>
<td>1.4</td>
<td>2.0</td>
</tr>
<tr>
<td>U Biscuits</td>
<td>-0.2</td>
<td>-1.2</td>
<td>-0.2</td>
<td>-2.0</td>
<td>1.5</td>
<td>-2.8</td>
<td>2.7</td>
</tr>
<tr>
<td>BT Telecom 1</td>
<td>-0.2</td>
<td>1.6</td>
<td>0.1</td>
<td>4.4</td>
<td>1.8</td>
<td>5.3</td>
<td>6.3</td>
</tr>
<tr>
<td>BT Telecom 2</td>
<td>0.8</td>
<td>-0.9</td>
<td>0.5</td>
<td>-12.6</td>
<td>-2.8</td>
<td>7.8</td>
<td>5.2</td>
</tr>
<tr>
<td>Wellcome</td>
<td>1.0</td>
<td>2.4</td>
<td>0.4</td>
<td>-1.9</td>
<td>6.2</td>
<td>3.7</td>
<td>7.6</td>
</tr>
<tr>
<td>Average</td>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
<td>-4.1</td>
<td>2.1</td>
<td>2.6</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Table 6.3.1b Relative performance for groups of options in the empirical and simulation study. (*values in italics)
Units: Percentage points of final cumulative value of cash inflow minus cash outflow divided by the notional value of the contract

<table>
<thead>
<tr>
<th>Group Variant</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>D*</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1E</td>
<td>0.1</td>
<td>-1.5</td>
<td>0.1</td>
<td>-8.2</td>
<td>-1.8</td>
<td>2.2</td>
<td>1.8</td>
</tr>
<tr>
<td>Group 1</td>
<td>0.2</td>
<td>-1.0</td>
<td>0.1</td>
<td>-1.0</td>
<td>4.9</td>
<td>6.7</td>
<td>6.5</td>
</tr>
<tr>
<td>Group 4E</td>
<td>2.1</td>
<td>4.5</td>
<td>0.7</td>
<td>1.3</td>
<td>8.2</td>
<td>4.6</td>
<td>9.6</td>
</tr>
<tr>
<td>Group 4</td>
<td>1.0</td>
<td>3.7</td>
<td>0.3</td>
<td>3.4</td>
<td>9.3</td>
<td>9.3</td>
<td>11.3</td>
</tr>
<tr>
<td>G4E - G1E</td>
<td>2.2</td>
<td>6.0</td>
<td>0.6</td>
<td>10.5</td>
<td>10.0</td>
<td>2.4</td>
<td>7.8</td>
</tr>
<tr>
<td>G4 - G1</td>
<td>0.8</td>
<td>4.7</td>
<td>0.2</td>
<td>4.4</td>
<td>4.4</td>
<td>2.6</td>
<td>4.8</td>
</tr>
</tbody>
</table>
6.3.1.1 Main results from Table 6.3.1a and Table 6.3.1b

In this section, we give the main results from Table 6.3.1a and Table 6.3.1b.

The final row of Table 6.3.1a indicates that the performance of all variants, except D, averaged over 30 options is higher than that of delta hedging, and that there are marked differences in relative performance for different variants. The columns are generally highly variable. This suggests that different variants may be appropriate for different groups of options, and as options change their moneyness, the appropriate variant to use may also change.

In Table 6.3.1b, we compare roughly the performance for, respectively, Group 1E and Group 4E, to that for Group 1 and Group 4. For Group 1E, as for Group 1 in the simulation, the performance of A and C is very close to DH; DH outperforms D, and no other difference is significant. For Group 4E, as for Group 4 in the simulation, D*, E and F outperform DH; but, unlike in Group 4, B outperforms DH. The differences for the other variants are also positive, although not significant. The differences between Group 4E and Group 1E are all positive, with largest differences for D and D*.

6.3.2 Relative performance between variants

In Table 6.3.2, we compare the relative performance between variants. Because the high correlation between B and D that was found in the simulation study is not found in this empirical study, we do not exclude any variants from this table.

<table>
<thead>
<tr>
<th>Variant</th>
<th>B-A</th>
<th>D-C</th>
<th>F-E</th>
<th>E-C</th>
<th>F-D</th>
<th>D*-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1E</td>
<td>1.2</td>
<td>4.2**</td>
<td>0.1</td>
<td>2.1</td>
<td>10.0</td>
<td>6.4</td>
</tr>
<tr>
<td>Group 1</td>
<td>-1.2</td>
<td>-1.1</td>
<td>-0.2</td>
<td>6.6**</td>
<td>6.6**</td>
<td>4.4**</td>
</tr>
<tr>
<td>Group 4E</td>
<td>0.6</td>
<td>0.6</td>
<td>5.0</td>
<td>3.9</td>
<td>8.3</td>
<td>6.9</td>
</tr>
<tr>
<td>Group 4</td>
<td>2.7</td>
<td>3.1</td>
<td>2.0</td>
<td>9.0</td>
<td>7.9</td>
<td>5.9</td>
</tr>
</tbody>
</table>

*(t-values in italics)*

Units: Percentage points of final cumulative value of cash inflow minus cash outflow divided by the notional value of the contract.
6.3.2.1 Main results from Table 6.3.2

In this section, we give the main results from Table 6.3.2.

For Group 1E, of the selected pairs of variants, C outperforms D, and F outperforms D; E-C and D*-D are both positive, but neither is significant. For Group 4E, of the three pairs of variants for Abrupt change - 95% level, only F-E is significantly different. In contrast the variants with transaction costs in their objective function outperform their counterpart variants, and D* outperforms D. These are broadly similar to the results from the simulation.

6.4 Statistical testing of hypothesis

In this section we test several of the hypotheses that were tested in the simulation study, and report on any difference in the state of corresponding hypotheses.

6.4.1 Hypotheses on relative performance and intergroup performance

In this section, the bases for the hypotheses are the same as those given above for the same hypotheses used in the simulation study (Sections 5.7.1 and 5.7.2). Of the hypotheses tested for the simulation study, H1, H4 and H8 are applicable to the empirical study; these are numbered H1E, H4E and H8E, and the results are given below, together with the corresponding results from the simulation study.

Table 6.4.1 Results of tests on relative performance and intergroup performance.

<table>
<thead>
<tr>
<th>(R=Reject) Hypothesis</th>
<th>D</th>
<th>D*</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1E DH will outperform the variants for Group 1E</td>
<td>Accept</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>H1 DH will outperform the variants for Group 1</td>
<td>Accept</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>H4E The variants will outperform DH for Group 4E</td>
<td>R</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H4 The variants will outperform DH for Group 4</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H8E G4E will outperform G1E</td>
<td>Accept</td>
<td>Accept</td>
<td>R</td>
<td>Accept</td>
</tr>
<tr>
<td>H8 G4 will outperform G1</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
</tbody>
</table>
6.4.1.1 Comments on the results of the tests

In general, the commentary in Section 5.7.1.1 on the results for H1 and H4 applies to H1E and H4E. For the high performing variants (D*, E and F) H1E is rejected, as was H1, and H4E is accepted, as was H4. The discussion on H1 and H4 applies equally to H1E and H4E. H8E is accepted for D, D*, and F; however it is rejected for E, whereas H8 was accepted for E. A possible reason for the rejection is that the differentiation between G4E and GIE is not sufficient to test H8E for the 95%-Level variant E, where abrupt changes are not critical to the performance of the variant.

6.4.2 Hypotheses on intervariant performance

The intervariant hypotheses tested in the simulation study for four groups are here tested for Groups 1E and 4E.

Table 6.4.2 Results of tests on intervariant performance.
(R=Reject, n.a.=not applicable)

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>G1E</th>
<th>G1</th>
<th>G4E</th>
<th>G4</th>
</tr>
</thead>
<tbody>
<tr>
<td>H11 E outperforms C</td>
<td>R</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H12 F outperforms D</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H13 F outperforms E for G4E</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H13.1 F does not outperform E for G1E</td>
<td>Accept</td>
<td>Accept</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>H14 D* outperforms D</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
</tbody>
</table>

6.4.2.1 Comments on the results of the tests

The intervariant hypotheses are accepted for G4E, as they were for G4, and with the exception of H11 they are accepted for G1E as they were for G1. H11 is rejected for G1E, with E not significantly outperforming C.
7 Conclusion

On the basis of the results of Section 5, with some support from the limited empirical study reported in Section 6, we have established that there are three high performing variants, E, F and D*, that significantly outperform DH for the options with several crossovers and abrupt changes, and also significantly, but to a lesser extent, outperform DH for options with very few, if any, crossovers and abrupt changes. In this sense the relative performance of the high performing variants is robust.

The results of the limited empirical study given in Section 6 are mainly consistent with the results from the simulation study. Despite the small number of options in the empirical groups, these results offer some support to the results from the simulation, especially in respect of the relevance of transaction costs, and to the characteristics of the time series of stock price for which the minimax variants perform best.
Appendix 1

Weighting hedge recommendations

Changes in stock price often contain an element of noise; the hedger may wish to react only to the signal element in the change. The probability that the hedger will consider that a given $\Delta S_t$ has a large noise element is higher, the higher the standard deviation of $S_t$. Some crossovers leave $S_t$ close to $X$; others leave $S_t$ far from $X$. The probability of another crossover occurring is higher the closer $S_t$ is to $X$, and the recommendation on $n_t$, the number of shares to hold, has a higher probability of being reversed, the closer $S_t$ is to $X$. The hedger is less likely to accept the transaction costs in realizing the recommended $n_t$ when he considers that another crossover is likely to occur, involving another possibly large and countervailing change in recommended $n_t$ with correspondingly large transaction costs. We give below an expression that the hedger can use to weight the number of shares recommended by a strategy to reflect his perception of the amount of noise in $\Delta S_t$ and the potential reversibility of the crossover. The hedger can give any weights to $a$ and $b$ to reflect the importance he attached to the standard deviation of $S_t$. The general expression is:

$$k_1 = \frac{w_1 \frac{\Delta S_t}{a \cdot sd} + w_2 \frac{S_t - X}{b \cdot sd}}{(w_1 + w_2)}$$

He may give a non-zero value to $w_2$ when there is a crossover from $t-1$ to $t$; there are no restrictions on the value he gives to $w_2$.

In Section 4.7, we apply the expression to minimax variant $D$, to give a weighted version of that variant, $D^*$. 

161
Appendix 2

Variation in the performance of minimax with transaction costs

In Section 4.6, we discussed transaction costs as part of Objective Function 2 of the minimax hedging strategy. Three variants constrain transaction costs: E and F have a transaction cost term (TC) in their objective function and D* uses a heuristic method. In this appendix, we present simulation results showing the variation in the performance of E and F based on two levels of the transaction cost $K$ and different values of $q$, the coefficient of the transaction cost term in Objective Function 2.

The two levels of $K$ are $K = 0.005$ and $K = 0.02$. Holding $K$ constant, we simulated the performance of E and F for different values of $q$. In the simulation, we used a constant volatility of 0.40. The simulation results are shown graphically in Figure A2, where the average relative performance for 250 options (corresponding to 250 replications) is plotted against $q$.

Figure A2 shows the variation in the relative performance of E and F with $q$ when $K = 0.005$ and when $K = 0.02$. The figure shows that there is an suitable value of $q$ for a given level of $K$ that results in the best relative performance for variants E and F. When $K = 0.02$, the suitable value of $q$ is 100, and when $K = 0.005$, the suitable value is 2000. When $q = 0$, the performance of E and F, respectively, is the same as that of variants C and D; the transaction cost term is effectively not used. When $q = 10^6$, the transaction cost term dominates the objective function and the hedge ratio is effectively unchanged during the life of the option.
Figure A2  Relative performance of variants E and F for different values of $q$, for $K=0.005$ and $K=0.02$.

Legend:

- $F_{02}$: Relative performance of variant F when $K=0.02$.
- $E_{02}$: Relative performance of variant E when $K=0.02$.
- $F_{005}$: Relative performance of variant F when $K=0.005$.
- $E_{005}$: Relative performance of variant E when $K=0.005$. 
5

Multi-Period Minimax Hedging Strategies

1. Introduction

In Chapter 4, we presented several variants of a minimax strategy, hereafter called Basic minimax, that determines the number of shares that minimizes the worst-case potential hedging error for the next period. In this chapter, we present two extensions to one of the three high performing variants, variant E. We consider extensions to this variant because, first, like variant F, it includes transaction costs in its objective function, and, second, unlike F, which may give rise to very extreme minimax hedging errors in a multi-period setting, E computes the minimax hedging error based on the most likely future values of the stock price, i.e. the 95%-Level worst case scenario.

The first extension is a two-period minimax strategy, hereafter called Two-Period minimax, where the worst case is defined over a two-period setting. In this extension, the objective function of Basic minimax is augmented to include the hedging error for the second period. In Basic minimax we have a one time period setting that is equal to the rebalancing interval. In Two-Period minimax we have a two time period setting, but the rebalancing interval remains equal to one period.

The second extension is a variable minimax strategy, hereafter called Variable minimax, where early rebalancing is triggered by the minimax hedging error. In Basic minimax, we preset a rebalancing interval that is constant throughout the life of the option and the
hedger necessarily rebalances at the end of each interval, and may not rebalance within that interval. In Variable minimax we also preset a constant rebalancing interval, but the hedger may rebalance before the end of that interval. Under Variable minimax the hedger can monitor the actual hedging error within a rebalancing interval; if he finds the actual hedging error unacceptable, he can rebalance before the end of that interval.

In Section 2, we present Two-Period minimax, and in Section 3, Variable minimax. In Section 4, we summarize the results of a simulation study where the performances of Basic minimax, the two multi-period extensions and delta hedging are compared. In Section 5, we illustrate the suitability of the multi-period strategies.

2 Two-Period Minimax Strategy

This strategy provides the hedger with a tool for computing the minimax hedging error in two time periods. We designed the strategy for the hedger who wishes to have a constant rebalancing interval, $\Delta t$. He decides on the number of shares to hold on the basis of the calculated minimax hedging error for the following two time periods, and he rebalances at the end of the first of the two periods. If he cannot rebalance at the end of a period because, say, there is a shortage of stock, by having taken into account the potential hedging error in the second period in deciding on the number of shares to hold at the start of the first period, he ensures that the actual hedging error in the second period will not be worse than the component of the Two-Period minimax hedging error for that period.

Two-Period minimax considers a worst case scenario for two time periods. The difference between this two-period worst case and the one-period worst case is that the range for the future stock price $S_{t+2}$ is wider for the two-period case. This results in a higher minimax hedging error for the two periods compared to the summation of the errors over two periods when using Basic minimax. In this sense, Two-Period minimax is a more cautious strategy than Basic minimax.

\footnote{The Two-Period minimax hedging error is given by Eqn (2.11).}
2.1 Minimax problem formulation

The minimax problem is given by

\[
\min_{n_t} \max_{(S_{t+1}, S_{t+2})} f(n_t, S_{t+1}, S_{t+2})
\]  

subject to

\[
S_{t+1} \leq S_{t+1} \leq S_{t+1} \quad (2.2)
\]

\[
S_{t+2} \leq S_{t+2} \leq S_{t+2} \n
\]

where \( f(n_t, S_{t+1}, S_{t+2}) \) is the objective function, presented in Section 2.2.

There are no constraints on \( n_t \), the number of shares to hold at time \( t \): non-negative \( n_t \) implies a long position in shares; negative \( n_t \) implies a short position in shares.

2.2 The objective function

We define \( U_1 : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k \), \( U_2 : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k \), \( U_3 : \mathbb{R}^k \to \mathbb{R}^k \), \( U_d : \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^{k+2} \), \( n_t \in \mathbb{R}^k \), \( S_{t+1} \in \mathbb{R}^k \), \( S_{t+2} \in \mathbb{R}^k \) and \( Q \) as a \((k+2) \times (k+2)\) positive definite diagonal weighting matrix. \( U_1 \) refers to the potential hedging error for the first time period and \( U_2 \) refers to that for the second time period. \( U_3 \) refers to a penalty term for transaction costs associated with buying or selling of stocks. \( U_d \in \mathbb{R}^{k+2} \) is the vector of desired values for the potential hedging error for the two periods and the transaction cost terms: we use a desired value of zero, i.e. the desired hedging error is zero and the desired transaction cost is zero.

The objective function is given by

\[
f(n_t, S_{t+1}, S_{t+2}) = \frac{1}{2} < U - U_d, Q(U - U_d) >
\]  

---

2See Chapter 4, Section 4.2.1.
3The hedger bought the shares.
4The hedger sold the shares.
5In minimax, we have the opportunity to adopt any desired value. We adopted a desired value of zero because in delta hedging, the expected value of the hedging error is zero.
where

\[
\begin{align*}
    n_i &= \begin{bmatrix}
        n^1_i \\
        \vdots \\
        n^k_i
    \end{bmatrix} \\
    S_{t+1} &= \begin{bmatrix}
        S^1_{t+1} \\
        \vdots \\
        S^k_{t+1}
    \end{bmatrix} \\
    S_{t+2} &= \begin{bmatrix}
        S^1_{t+2} \\
        \vdots \\
        S^k_{t+2}
    \end{bmatrix}
\end{align*}
\] (2.4)

\[
U(n_i, S_{t+1}, S_{t+2}) = \begin{bmatrix}
    U_1(n_i, S_{t+1}) \\
    \vdots \\
    U_{3}(n_i)
\end{bmatrix}
\] and

\[
U^d = \begin{bmatrix}
    U^d_1 \\
    \vdots \\
    U^d_3
\end{bmatrix}
\] (2.5)

\[
U_1(n_i, S_{t+1}) = \sum_{i=1}^{k} n^i_i (S^i_{t+1} - S^i_t) + \sum_{i=1}^{k} N^i (B^i_t - B^i_{t+1}(S^i_{t+1})) + \sum_{i=1}^{k} (-n^i_i - n^i_{i-1}) S^i_t + C^i_{t-1} (1 + r \Delta t) r \Delta t
\] (2.6)

\[
U_2\left(n_i, \begin{bmatrix} S^i_{t+1} \\ S^i_{t+2} \end{bmatrix}\right) = \sum_{i=1}^{k} n^i_i (S^i_{t+2} - S^i_{t+1}) + \sum_{i=1}^{k} N^i (B^i_t - B^i_{t+1}(S^i_{t+1} - B^i_{t+1}(S^i_{t+2})))
\]

\[
+ \sum_{i=1}^{k} (-n^i_i - n^i_{i-1}) S^i_t + C^i_{t-1} (1 + r \Delta t) (1 + r \Delta t) (r \Delta t)
\] (2.7)

where

\[
C^i_{t-1} = C^i_{t-2} (1 + r \Delta t) - (n^i_{t-1} - n^i_{t-2}) S^i_{t-1} - \hat{K}(n^i_{t-1} - n^i_{t-2}) S^i_{t-1}
\] (2.8)

\[
U_3(n_i) = \begin{bmatrix}
    U^1_3(n^1_i) \\
    \vdots \\
    U^k_3(n^k_i)
\end{bmatrix}
\] (2.9)

where

\[
U^i_3(n^i_i) = \hat{K}(n^i_i - n^i_{i-1}) S^i_t
\] (2.10)

We first identify all the variables in Eqns (2.3) to (2.10) and then give an economic interpretation of Eqns (2.6) and (2.7).

\(C^i_{t-1}\) is the cumulative value of \textbf{cash inflow minus cash outflow} at time \(t-1\). \(C^i_{t-1}(1 + r \Delta t)\) is \(C^i_{t-1}\) with interest payments. The first term of (2.8) is the cumulative value of cash inflow minus cash outflow from the previous period with interest payments.

\(\hat{K}\) Eqn (2.10) is discussed in detail in Chapter 4, Section 4.6.
The second term is a cash outflow if the \( n_{t-1}^i > n_{t-2}^i \); otherwise, it is a cash inflow. The third term is always a cash outflow. We note that \( C_{t-1}^i \) will normally be a negative number. At time \( t \), \( C_{t-1}^i \) is a constant: all the variables in (2.8) have actual values.

Because transaction costs\(^7\) introduce nondifferentiability into the equation, they do not come into the objective function as part of \( U_1 \) nor \( U_2 \). Instead, we introduce \( U_3^i \) to represent a penalty for transaction costs for each option \( i \) at time \( t \).

In the weighting matrix \( Q \), the weights \( q_i, \ i = 1, \ldots, k + 2 \), which are specified by the hedger, represent his preferences: a high \( q_1 \) represents an emphasis on minimizing the potential hedging error that may be incurred from time \( t \) to time \( t + 1 \); a high \( q_2 \) represents an emphasis on minimizing the potential hedging error that may be incurred from time \( t + 1 \) to time \( t + 2 \). High \( q_i, \ i = 3, \ldots, k + 2 \), represents an emphasis on minimizing the corresponding transaction cost term.

For each option \( i, i = 1, \ldots, k \), \( B_{t+1}^i(S_{t+1}^i) \) and \( B_{t+2}^i(S_{t+2}^i) \) is determined using the Black and Scholes\([1]\) option pricing model\(^8\) and using Leland's\([20]\) modified volatility estimate\(^9\).

We now give an economic interpretation of Eqns (2.6) and (2.7). \( U_1 \) represents the potential hedging error between time \( t \) and time \( t + 1 \); it comprises the potential shift in the stock position, the potential shift in the option position and the potential interest payment. \( U_2 \) represents the potential hedging error between time \( t + 1 \) and time \( t + 2 \). \( U_3 \) refers to a penalty term for transaction costs associated with buying or selling of stocks. We wish to minimize the potential hedging error, including interest payments on borrowed money, for two time periods. At the same time, we wish to constrain transaction costs.

\(^7\)The treatment of transaction costs is discussed in detail Chapter 4, Section 4.6.
\(^8\)Eqns (4.12), (4.13) and (4.14) in Chapter 4, Section 4.4.2.1.
\(^9\)Eqn (3.15) in Chapter 4, Section 3.2.
2.3 The Two-Period minimax hedging error

In contrast to the minimax hedging error for one time period given by Eqn (4.23) in Section 4.5 of Chapter 4, the minimax hedging error for two time periods is given by,

\[
\text{Two-Period minimax hedging error} = U_1(n, S_{i+1}) + U_2(n, S_{1, i}, S_{2, i}) \quad (2.11)
\]

3 Variable Minimax Strategy

This strategy is the same as Basic minimax in all respects except that under Variable minimax the hedger can rebalance within the preset interval, and in deciding when to rebalance he takes account of the actual hedging error. If the actual hedging error is unacceptable to him, he may wish to rebalance before the end of the preset interval; if the actual hedging error is not unacceptable, he would rebalance at the end of the preset interval.

At the start of each time period, the hedger specifies his worst case scenario for that time period. One such scenario may be a large movement in the price of the underlying stock that results in an actual hedging error that is unacceptable to him. He uses Basic minimax to minimize the potential hedging error that corresponds to such a scenario. If the stock price moves in the direction that makes the hedging error unacceptable to him, he may wish to rebalance early. For example, if the preset rebalancing interval is 1 week, i.e. 5 trading days, he may rebalance on day 1, 2, 3 or 4, and the next preset interval will start on the following day.

Under Variable minimax, the hedger could use the minimax hedging error calculated\(^\text{10}\) by Basic minimax as a criterion in deciding whether to rebalance before the end of the preset interval. Because the minimax hedging error corresponds to the worst case scenario for the next period, if the stock price is within the preset range, he knows that the absolute

\(^\text{10}\)This is given by Eqn (4.23) in Chapter 4, Section 4.5.
value of the actual hedging error would not be higher than the minimax hedging error. Despite this knowledge, he may consider such an actual hedging error to be unacceptable to him. If he does, he can avoid the accumulation of unacceptable hedging errors by rebalancing early should the actual hedging error be worse than the threshold error, which is defined as the proportion of the minimax hedging error that is acceptable to him. This threshold error serves as a trigger for early rebalancing.

If, at any time within the preset interval, the hedger finds that the actual hedging error is worse than his threshold error, he may rebalance at that time. However, if, at any time within the preset interval, the actual error is not worse than his threshold error, he would not rebalance before the end of the preset interval. He can use the minimax hedging error as a criterion for deciding when to rebalance. The time at which he rebalances becomes the start of the next preset interval.

Because Variable minimax uses a system to monitor the actual hedging error and allows the hedger to rebalance early when the actual hedging error becomes unacceptable to him, Variable minimax is more responsive to unfavorable stock price movements than Basic minimax. In this sense, Variable minimax is a more aggressive strategy than Basic minimax.

### 3.1 Minimax problem formulation

The minimax problem is given by

\[
\min_{n, S_{t+1}} \max_{t} f(n_t, S_{t+1}) \tag{3.1}
\]

subject to \( S_t \leq S_{t+1} \leq S_u \) \tag{3.2}

where \( f(n_t, S_{t+1}) \) is the objective function, presented in Section 3.2, and \( S_t \leq S_{t+1} \leq S_u \) is the range defined under Worst Case 1, the 95%-Level\(^\text{11}\).
There are no constraints on $n_t$, the number of shares to hold at time $t$: non-negative $n_t$ implies a long position in shares\(^{12}\); negative $n_t$ implies a short position in shares\(^{13}\).

### 3.2 The objective function

We define $U_1: \mathbb{R}^k \times \mathbb{R}^t \to \mathbb{R}^t$, $U_2: \mathbb{R}^k \to \mathbb{R}^t$, $U: \mathbb{R}^k \times \mathbb{R}^t \to \mathbb{R}^{k+1}$, $n_t \in \mathbb{R}^t$, $S_{t+1} \in \mathbb{R}^t$ and $Q$ as a $(k+1) \times (k+1)$ positive definite diagonal weighting matrix. $U^d \in \mathbb{R}^{k+1}$ is the vector of desired values for the potential hedging error and the transaction cost terms: we use a desired value of zero, i.e. the desired hedging error is zero\(^{14}\) and the desired transaction cost is zero.

The objective function is given by

$$f(n_t, S_{t+1}) = \frac{1}{2} < U - U^d, Q(U - U^d) >$$

where

$$n_t = \begin{bmatrix} n_t^1 \\ \vdots \\ n_t^k \end{bmatrix} \quad \text{and} \quad S_{t+1} = \begin{bmatrix} S_{t+1}^1 \\ \vdots \\ S_{t+1}^k \end{bmatrix}$$

$U(n_t, S_{t+1}) = \begin{bmatrix} U_1(n_t, S_{t+1}) \\ \vdots \\ U_2(n_t) \end{bmatrix}$ and $U^d = \begin{bmatrix} U_{1d}^1 \\ \vdots \\ U_{2d}^k \end{bmatrix}$

$$U_1(n_t, S_{t+1}) = \sum_{i=1}^{k} n_t^i (S_{t+1}^i - S_t^i) + \sum_{i=1}^{k} N^i (B^i_t - B^i_{t+1}(S_{t+1}^i)) + \sum_{i=1}^{k} \left((n_t^i - n_{t-1}^i)S_t^i + C_{t-1}^i (1 + r\Delta t)\right) r\Delta t$$

where

$$C_{t-1}^i = C_{t-2}^i (1 + r\Delta t) - (n_{t-1}^i - n_{t-2}^i)S_t^i - \hat{K} (n_{t-1}^i - n_{t-2}^i)S_{t-1}^i.$$

\(^{12}\)The hedger bought the shares.

\(^{13}\)The hedger sold the shares.

\(^{14}\)In minimax, we can adopt any desired value. We adopted a desired value of zero because in delta hedging, the expected value of the hedging error is zero.
\[ U_2(n_i) = \begin{bmatrix} U_2^1(n_i^1) \\ \vdots \\ U_2^m(n_i^m) \end{bmatrix} \]  

(3.8)

where\(^{15}\) \[ U_2^i(n_i^i) = \hat{K}(n_i^i - n_{i-1}^i)S_i^i. \]  

(3.9)

Eqns (3.3) to (3.9) constitute the Basic minimax formulation presented in Chapter 4, corresponding to Eqns (4.16) to (4.22). The definition of variables and economic interpretation given in Chapter 4 apply. The minimax hedging error\(^{16}\) is \( U_1(n_i, S_{i+1}) \). Variable minimax is Basic minimax augmented by a system to monitor the actual hedging error.

3.3 The monitoring system

The minimax potential hedging error, with the corresponding minimizing variable \( n_i \), applies to the time period between \( t \) and \( t + 1 \); hereafter we refer to this time period as \( \tau \).

We define a smaller interval \( \Delta t \) such that \( m\Delta t = \tau \) where \( m \) is the number of small intervals of length \( \Delta t \). Here, we consider multi-periods within the maximum period, \( \tau \).

On solving the minimax problem, we find the value of \( n_i \) that minimizes the maximum hedging error that could occur within the time period \( t \) to \( t + 1 \), given the preset range of \( S_{i+1}, \) i.e. the maximum hedging error is the value that we have insured against, when using \( n_i \). It should be noted that there may be combinations of time \( t + m_0 \Delta t, \) \( m_0 = 0, 1, \ldots, m \), and corresponding stock price \( S_{i+m_0\Delta t} \), that give the same hedging error as the one defined by the minimax solution.

For each small interval \( \Delta t \), we define a certain percentage \( x\% \) of the absolute value of the minimax hedging error \( M \) as the threshold \( V \), i.e.

\[ V = \frac{x}{100} M. \]  

(3.10)

\(^{15}\)Eqn (3.9) is discussed in detail in Chapter 4, Section 4.6.

\(^{16}\)See Chapter 4, Section 4.5.
For each time period $t$ to $t + m_0 \Delta t$, $m_0 = 0, 1, \ldots, m$, we calculate the absolute value of the actual hedging error, $A$, and compare this with the threshold $V$. If the actual hedging error is negative and

$$A_{t + m_0 \Delta t} \geq V$$

at time $t + m_0 \Delta t$, with actual stock price $S_{t + m_0 \Delta t}$, then we rebalance and solve the minimax problem again, and update $t$, i.e., set to $t + m_0 \Delta t$ (the current time). If condition (3.11) is not satisfied for any time $t + m_0 \Delta t$, $m_0 = 0, 1, \ldots, m$, then we rebalance at time $t + m \Delta t$.

4 Simulation study of the performance of different strategies

In this section, we simulate the relative performance of the multi-period minimax strategies when they are used to hedge the risk of writing a European call option. We use the definition of performance and relative performance given in Section 5.6 of Chapter 4. This simulation consisted of 1250 replications, subdivided into 5 levels of sigma.

4.1 Objective of the simulation

This simulation is intended to serve as a feasibility study on potential extensions to Basic minimax. Towards this, the simulation is used to ascertain whether the two multi-period extensions of Basic minimax outperform delta hedging (DH), to ascertain whether Basic minimax outperforms Two-Period minimax, which is designed to be a more cautious strategy than Basic minimax, and to ascertain whether Variable minimax, which is designed to be a more aggressive strategy than Basic minimax, outperforms Basic minimax.
4.2 Generation of data
The method for generating stock price series and option price series is described in Chapter 4, Section 5.2.

4.3 From set-up to wind-down
In Chapter 4 Section 5.4, we discussed the mechanics of hedging, from setting up to winding down the hedge. For this simulation, we use the same mechanics with a few changes in parameters. Delta hedging, Basic minimax, and Two-Period minimax involve rebalancing the hedge at uniform intervals of time; in this simulation, the interval is one week. Variable minimax involves the monitoring of the actual hedging error on a daily basis. Weekly and daily data include the price of the stock, the price of the option, the risk-free interest rate, the time to maturity and the volatility (sigma) of returns on the stock, given as one of five preset levels: 0.2, 0.3, 0.4, 0.5, and 0.6. The risk-free interest rate is preset at 0.10. Dividends are excluded from the analysis. In this simulation, \( N \), the number of contracted shares, is 100. In addition, for Variable minimax, we set the threshold level at 10% of the minimax hedging error.

4.4 Stratification
We applied delta hedging, Basic minimax, Two-Period minimax and Variable minimax to 1250 options; as in Chapter 4, these were subdivided into 5 levels of sigma: 0.2, 0.3, 0.4, 0.5 and 0.6. However, in contrast to the simulation in Chapter 4, we did not stratify in terms of groups because, as mentioned in Section 4.1, this simulation study is intended only to explore the feasibility of extensions to Basic minimax.

4.5 Results of the simulation study
All four strategies were used to hedge the risk of writing each of the 1250 European call options. The average performance in using a strategy is calculated for all options with a constant sigma level. In Table 4.5a, we summarize the relative performance of Basic minimax and the two multi-period extensions. In Table 4.5b, we summarize the
difference in relative performance between minimax strategies. These tables also contain, in italics in small font, the absolute value of the t-statistics, followed by ** if the difference is significant at the 2% level, and * if it is significant at the 10% level.

Table 4.5a Relative performance of three minimax strategies. (*-values in italics)

Units: Percentage points of final cumulative value of cash inflow minus cash outflow divided by the notional value of the contract

<table>
<thead>
<tr>
<th>Sigma</th>
<th>STRATEGY</th>
<th>Basic</th>
<th>Two-Period</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td></td>
<td>3.3</td>
<td>3.3</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td></td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td>2.9</td>
<td>2.9</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>2.3</td>
<td></td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td>2.9</td>
<td>2.4</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>2.2</td>
<td></td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td>2.8</td>
<td>1.5</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>2.1</td>
<td></td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td>2.8</td>
<td>1.7</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td></td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>2.9</td>
<td>2.3</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Table 4.5b Difference in relative performance between minimax strategies. (*-values in italics)

Units: Percentage points of final cumulative value of cash inflow minus cash outflow divided by the notional value of the contract

<table>
<thead>
<tr>
<th>Sigma</th>
<th>STRATEGY</th>
<th>Basic minus Two-Period</th>
<th>Variable minus Basic</th>
<th>Variable minus Two-Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td></td>
<td>0.0</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td>0.0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td>0.5</td>
<td>0.3</td>
<td>0.8</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td>1.3</td>
<td>1.1</td>
<td>1.4</td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td>1.1</td>
<td>0.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.6</td>
<td>0.2</td>
<td>0.8</td>
</tr>
</tbody>
</table>

17 We use significance levels of 2% and 10% to express the results of a 2-tailed test; however, we use a 1-tailed test for the sign of the difference at the 5% level.
4.5.1 Main results from Table 4.5a and Table 4.5b

In this section we give the main results from Table 4.5a and Table 4.5b.

From Table 4.5a, each strategy outperforms DH by about 3 percentage points, with Variable Minimax being slightly the better performer. Two-Period minimax is the worst performer, outperforming DH by just over 2 percentage points. For all three strategies, their relative performances fall with increasing sigma; the fall is most marked in Two-Period minimax.

From Table 4.5b, for low levels of sigma, Variable minimax outperforms Basic minimax, but for high levels of sigma, Variable minimax performs much the same as Basic minimax. For low levels of sigma, Two-Period minimax performs much the same as Basic minimax, but for high levels of sigma, Basic minimax outperforms Two-Period minimax.

4.6 Statistical testing of hypotheses

In this section we test a number of hypotheses on the difference in relative performance of the three minimax strategies: Basic, Two-Period and Variable. In Table 4.6, we list the hypotheses and give their accept/reject state at the 5% level of significance.\(^\text{18}\)

In Chapter 4, we have established that the minimax hedging strategy outperforms DH for a number of variants. Further, we have established that variant E, the variant that we use as Basic minimax, outperforms DH. Because Two-Period minimax has the same formulation as Basic minimax except for the inclusion of a second-period variable that determines the contribution of the second period to the worst-case scenario (see Section 2), we hypothesize (H1) that Two-Period minimax will outperform DH. Similarly, because Variable minimax has exactly the same formulation as Basic minimax except for

\(^{18}\)We use the 5% level of significance to reflect a less stringent criterion on the acceptance or rejection of a hypothesis because the simulation data exhibit a wide range of moneyness and the hypotheses we tested are not sufficiently refined to accommodate variations with respect to moneyness.
the inclusion of a monitoring system that determines an option to rebalance early, we hypothesize (H2) that Variable minimax will outperform DH.

Because Two-Period minimax considers the worst case scenario for two time periods, it computes the minimax hedging error for two periods (Eqn (2.11)) based on a wider range for the second period compared to the range in Basic minimax when Basic minimax is used for the second period. Because Two-Period minimax computes its hedge recommendation based on the contribution of the second period to the minimax hedging error, the hedge recommendation is not necessarily consistent with the hedge recommendation of Basic minimax, and the hedge recommendation of Two-Period minimax may result in a worse hedging error if the worst case occurs during the first period. Given the reason above and because the hedger follows a policy of rebalancing after one period, we hypothesize (H3) that Basic minimax outperforms Two-Period minimax.

Because in Variable minimax the hedger can avoid incurring large negative hedging errors, we hypothesize (H4) that Variable minimax outperforms Basic minimax.

If Hypotheses H3 and H4 are accepted, then by transitivity, Variable minimax outperforms Two-Period minimax. In case H3 and H4 are not accepted for some sigma, we hypothesize (H5) that Variable minimax outperforms Two-Period minimax.
Table 4.6 The accept/reject state of the hypotheses\textsuperscript{19}, for each level of sigma. (\(R=\text{Reject}\))

<table>
<thead>
<tr>
<th>Hypothesis:</th>
<th>Sigma:</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 Two-Period outperforms DH</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H2 Variable outperforms DH</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H3 Basic outperforms Two-Period</td>
<td>(R)</td>
<td>(R)</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H4 Variable outperforms Basic</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>(R)</td>
<td>(R)</td>
<td>Accept</td>
</tr>
<tr>
<td>H5 Variable outperforms Two-Period</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
<td>Accept</td>
</tr>
</tbody>
</table>

\textbf{4.6.1 Comment on the rejected hypotheses}

H3 was rejected for low levels of sigma. A possible reason for this may be that when sigma is low the ranges defined by the 95%-Level are relatively small. For Two-Period minimax, the range for the second period may not be large compared to the range for Basic minimax when it is used for the second period. This may have resulted in similar minimax hedging errors under Two-Period minimax and Basic minimax.

H4 was accepted for low levels of sigma but rejected for higher levels. This may suggest that the relatively large 95%-Level ranges at higher levels of sigma resulted in very high minimax hedging errors that gave high threshold errors that were rarely exceeded. If the threshold errors are not exceeded, Variable minimax performs much the same as Basic minimax. This suggests that the percentage to use in calculating the threshold value is critical to the performance of Variable minimax and should be made a function of sigma.

\textsuperscript{19}Except for Hypothesis 5, we tested the hypotheses by defining the Null Hypothesis as no significant difference between means; the Alternative Hypothesis corresponds to the hypothesis outlined above. In Table 4.6, an accept state means an acceptance of the Alternative Hypothesis.
4.7 Rank ordering

The simulation results suggest the following rank order of positive differences in performance. Table 4.7 gives the rank ordering of the strategies.

Table 4.7  Rank order of positive significant differences in performance for each level of sigma.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Sigma level</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable Minimax</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Basic Minimax</td>
<td>2.5</td>
<td>2.5</td>
<td>2</td>
<td>2</td>
<td>1.5</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>Two-period minimax</td>
<td>2.5</td>
<td>2.5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>Delta Hedging</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4.0</td>
<td></td>
</tr>
</tbody>
</table>

All minimax strategies for the 5 levels of sigma outperform delta hedging. For low levels of sigma (sigma=0.2 or 0.3), Basic minimax has the same rank as Two-Period minimax while for a high level of sigma (sigma=0.6), Basic minimax has the same rank as Variable minimax.

For the average rank order, Variable minimax outperforms Basic minimax, which is consistent with the view that Variable minimax is more responsive to the development of unacceptable hedging errors. Basic minimax outperforms Two-Period minimax, which is consistent with the view that Two-Period minimax is less suitable when the hedger rebalances at the end of one period.
5 Illustrations

In this section, we present illustrations showing the performance of the strategies. In Section 5.1, we illustrate the performance of Variable minimax and in Section 5.2, that of Two-Period minimax.

5.1 Variable minimax: an illustration

In this section, we illustrate the performance of Variable minimax depending on how the state of the option changes over time; in particular we show the performance of an option that is:

1. progressively deep-in-the-money;
2. remains close to at-the-money;
3. progressively deep-out-of-the-money.

The three time series of stock prices corresponding to the options described above are shown in Figures 5.1.1a, 5.1.2a and 5.1.3a. The exercise price of the option is represented by the horizontal line (Xprice = 1000). These figures are accompanied respectively by Figures 5.1.1b, 5.1.2b and 5.1.3b showing a graph of the minimax hedging errors and actual hedging errors. The shaded area gives the minimax hedging error at every time period, i.e. 5 trading days. The line graph gives the actual hedging error on a daily basis. The arrow in the figure points to the event that early rebalancing has been triggered in Variable minimax.
5.1.1 Performance of strategies when the option is progressively deep-in-the-money

Where the option is progressively deep-in-the-money, in general, the minimax strategies become increasingly unsuitable over time. Because the worst case scenario defined by the 95% confidence interval about the expected future stock price is not likely to cover the exercise price, there is a very high probability that the option will remain in-the-money for the next time period. When the option is sufficiently deep-in-the-money, a marginal increase in the stock price is equal to a marginal increase in the value of the option, and as time evolves, the choice of $n_t$ is driven towards $N$, the number of contracted shares. This implies that minimax is increasingly insensitive to stock price movements and that the minimax hedging error is increasingly a poor measure of the bound on the actual hedging error. This can be seen in Figure 5.1.1b where the minimax hedging error does not vary much despite the fact that the option is nearing maturity.

Figure 5.1.1a. The stock price series of an option that is progressively deep-in-the-money.

Figure 5.1.1b. The minimax hedging errors and actual hedging errors, where the arrow points to the actual hedging error that triggered an early rebalancing.
5.1.2 Performance of strategies when the option remains close to at-the-money

Where the option remains close to at-the-money, in general, the minimax strategies are suitable because the worst case scenario defined by the 95% confidence interval about the expected future stock price is likely to cover the exercise price; the possible oscillation of the stock price about the exercise price implies that there may be a high variation in potential hedging error corresponding to the 95%-Level. Because minimax is designed to be sensitive to this variation in potential hedging error, the minimax hedging error provides a good measure of the bound on actual hedging error. This can be seen in Figure 5.1.2b where the minimax hedging error (shaded area) increases in time as the option gets closer to maturity, anticipating an increase in the actual hedging error. We note that this increase in the minimax hedging error does not imply that minimax is increasingly unsuitable as a hedging strategy; it implies that minimax is increasingly sensitive to increasing variation in actual hedging error as time gets closer to maturity and the option remains close to at-the-money.

Figure 5.1.2a. The stock price series of an option that remains close-to-at-the-money.

Figure 5.1.2b. The minimax hedging errors and actual hedging errors, where the arrow points to the actual hedging error that triggered an early rebalancing.
5.1.3 Performance of strategies when the option is progressively deep-out-of-the-money

Where the option is progressively deep-out-of-the-money, in general, the minimax strategies become increasingly unsuitable over time. Because the worst case scenario defined by the 95% confidence interval about the expected future stock price is not likely to cover the exercise price, there is a very high probability that the option will remain out-of-the-money for the next time period. Because the value of the option reduces to zero over the life of the option, the choice of \( n \) is driven towards zero. This implies that minimax is increasingly insensitive to stock price movements and that the minimax hedging error is increasingly a poor measure of the bound on the actual hedging error. This can be seen in Figure 5.1.3b where the minimax hedging error does not vary much despite the fact that the option is nearing maturity.

Figure 5.1.3a. The stock price series of an option that is progressively-out-the-money.

Figure 5.1.3b. The minimax hedging errors and actual hedging errors, where the arrow points to the actual hedging error that triggered an early rebalancing.
5.2 Two-Period minimax: an illustration

We present a simple illustration where Two-Period minimax performed better than Basic minimax when the hedger failed to rebalance after the first time period.

At time 0, we have the following data, with \( N = 100 \).

<table>
<thead>
<tr>
<th>Strategy</th>
<th>( S_0 )</th>
<th>( X )</th>
<th>Option Price</th>
<th>Hedge Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta hedging</td>
<td>489</td>
<td>460</td>
<td>45</td>
<td>83</td>
</tr>
<tr>
<td>Basic minimax</td>
<td>489</td>
<td>460</td>
<td>45</td>
<td>66</td>
</tr>
<tr>
<td>Two-Period minimax</td>
<td>489</td>
<td>460</td>
<td>45</td>
<td>72</td>
</tr>
</tbody>
</table>

The minimax hedging error under Basic minimax for one time period is 636, based on \( 465 \leq S_{t+1} \leq 517 \), and that for Two-Period minimax (for two time periods) is 1024, based on \( 465 \leq S_{t+1} \leq 517 \) and \( 457 \leq S_{t+2} \leq 531 \).

From the data above, we see that the option is in the money. The hedge recommendation under Basic minimax suggests that the strategy found a solution consistent with a fall in the future stock price. The hedge recommendation under Two-Period minimax suggests that the strategy is more cautious. It anticipated that a fall in the stock price will result in the worst case hedging error for the first period, just like Basic minimax, but, in its calculation, it also considered a countervailing rise in the stock price for the second period. This way of searching for a solution is reflected in its hedge recommendation which is somewhat in between the hedge recommendation of delta hedging and that of Basic minimax.

At time 1, if the hedger was able to rebalance, his hedging error, under each strategy, is given below.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>( S_1 )</th>
<th>( X )</th>
<th>Option Price</th>
<th>Actual hedging error at time ( t+1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta hedging</td>
<td>484</td>
<td>460</td>
<td>41</td>
<td>-15</td>
</tr>
<tr>
<td>Basic minimax</td>
<td>484</td>
<td>460</td>
<td>41</td>
<td>70</td>
</tr>
<tr>
<td>Two-Period minimax</td>
<td>484</td>
<td>460</td>
<td>41</td>
<td>40</td>
</tr>
</tbody>
</table>

If the hedger failed to rebalance at time 1, and at time 2 the stock price did rise to a very high level, say the upper bound on \( S_2 = 531 \), as anticipated by Two-Period minimax, the actual hedging error for two time periods, under each strategy, is given below.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>( S_2 )</th>
<th>( X )</th>
<th>Option Price</th>
<th>Actual hedging error at time ( t+2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta hedging</td>
<td>531</td>
<td>460</td>
<td>70</td>
<td>1486</td>
</tr>
<tr>
<td>Basic minimax</td>
<td>531</td>
<td>460</td>
<td>70</td>
<td>772</td>
</tr>
<tr>
<td>Two-Period minimax</td>
<td>531</td>
<td>460</td>
<td>70</td>
<td>1024</td>
</tr>
</tbody>
</table>

The actual hedging error under Basic minimax exceeded its minimax hedging error. The cautious hedge recommendation under Two-Period minimax resulted in a more positive, and at the same time acceptable, hedging error than that under Basic minimax. As it turned out, delta hedging had the highest hedging error in absolute terms for the second period; this is due to a high hedge recommendation at time 0.
6 Conclusion

We have presented two multi-period extensions of the basic minimax hedging strategy that the hedger can use under specific situations. Two-Period minimax is designed for the hedger who wishes to consider the possibility that he may fail to rebalance at the end of a preset, one-period, interval. Variable minimax is designed for the hedger who wishes to actively avoid negative hedging errors. The results of the simulation suggest the following rank ordering of the strategies: Variable minimax, Basic minimax, Two-Period minimax, delta hedging.

Variable minimax is a more aggressive strategy in the sense that the hedger regularly monitors the development of the actual hedging error and rebalances early in response to an undesirable event. The strength of Variable minimax is that it gives the hedger a criterion that provides him with the opportunity to rebalance early in order to limit his accumulated hedging errors to some acceptable level. However to the extent that more frequent rebalancing increases transaction costs, any benefits from early rebalancing may be offset by such an increase, even though the Variable minimax constrains these costs.

Two-Period minimax is a more cautious strategy and, as such, it is less suitable than Basic minimax when the hedger can rebalance at the end of the first period. The strength of the Two-Period minimax is that it provides the hedger with a buffer that can, to some extent, absorb the negative effects of an undesirable event that may occur during the second period, if he failed to rebalance at the end of the first period.

The two multi-period extensions studied in this chapter have been shown to perform well in the circumstances they have been designed for. The results of the simulation are such that further advances would be expected through studies of Variable minimax under different degrees of moneyness, and studies of Two-Period minimax under changing levels of volatility. However we do not consider that Variable minimax could usefully be extended to an \( n \)-period setting.
1 Introduction

In Chapter 4, we presented several variants of a minimax strategy, hereafter called Basic minimax, that determines for an individual option the number of shares that minimizes the worst-case potential hedging error for the next period. In Section 4.2 of that chapter we defined the worst case scenario in terms of movements in the price of the underlying stock, with the source of uncertainty being the total volatility of returns on the underlying stock. In this chapter, we define the worst case scenario in terms of the two components of total volatility: the market risk and the specific risk. We use the Capital Asset Pricing Model (CAPM) as basis for a price determination function, and develop a CAPM-based minimax hedging strategy that uses both components of total volatility; we refer to this strategy as CAPM minimax.

In this chapter, we consider the problem of hedging the risk of writing a portfolio of call options. This problem can be restated as the problem of hedging the risk of holding a portfolio of written call options. The risk of holding such a portfolio can be hedged by holding a portfolio of underlying stocks. We define the hedge portfolio as the combination of the portfolio of written call options and the portfolio of the underlying stocks.
We developed CAPM minimax because, as we show in Section 3, when Basic minimax is used to hedge the risk of holding a portfolio of more than one written call option, its performance is inferior to its performance when it is used to hedge the risk of writing the options individually. This inferior performance of Basic minimax when it is applied to a portfolio may be caused by a compounding of the worst case scenarios the hedger has specified for each option, giving a worst case scenario for the portfolio that may markedly overstate his view of the worst case for the portfolio. The hedger may consider that this compounded scenario has a very low probability of being realized, and so may choose a less severe worst case scenario for the portfolio.

As we noted above, CAPM minimax uses both components of total volatility, and so it can incorporate information about these components separately in its search for a solution. For an individual written call option, we express the worst case scenario in terms of total volatility because we wish to define a range for future stock price movements; in this case it is not necessary to separate total volatility into its components to define the range. In contrast, for a portfolio of several written call options, we express the worst case scenario in terms of the components of total volatility for the following reason. As the number of written options increases, the specific risk of the portfolio of underlying stocks decreases. At the limit, where options are written for all the stocks in the market, the specific risk of the portfolio of underlying stocks reaches a minimum value. For such a portfolio of underlying stocks, where specific risk may be very small compared to the market risk, the remaining source of uncertainty for the value of this portfolio becomes the market risk. Further, for a given number of underlying stocks in a portfolio, this portfolio's market risk will be more dominant than its specific risk if the underlying stocks in the portfolio have high market risks and low specific risks; equally, the portfolio's market risk will be less dominant if the underlying stocks have low market risks and high specific risks. Where the portfolio's market risk is more dominant than its specific risk, the compounded scenario of individual worst cases based on total volatility would
overstate the worst case scenario the hedger would specify if he based his specification on market risk.

In Section 2, we present the Capital Asset Pricing Model, the Market Model, CAPM minimax and the definitions of two worst case scenarios. In Section 3, we present the results from a simulation when two variants of Basic minimax and of CAPM minimax are applied to 150 portfolios of 5 options. In Section 4, we present the beta-risk profile of a hedge portfolio under CAPM minimax.

2. The CAPM-based minimax hedging strategy

In this section, we develop the CAPM-based minimax hedging strategy using a price determination function derived from the Market Model. We first present the Capital Asset Pricing Model, followed by the formulation of CAPM minimax and its worst case scenarios.

2.1 The Capital Asset Pricing Model

In this section we summarize the Capital Asset Pricing Model (CAPM) and develop from it a price determination function.

Let: \( \beta \) beta of the stock  
\( R_S \) total return on the stock  
\( R_M \) return on the market index  
\( \varepsilon_S \) disturbance variable  
\( r_f \) risk-free rate  
\( S_t \) stock price at time \( t \)  
\( M_t \) index level at time \( t \)
The CAPM is an expectations theory where the expected return on a stock is a function of the expected return on the market. This is given by

\[ E(R_{s,t}) = r_f + \beta (E(R_{\hat{M},t}) - r_f) \]  

(2.1)

where \( \hat{M} \) is the market portfolio. We assume that there is a market proxy, represented by an index \( M \), having a beta equal to 1. The Market Model\(^2\), based on the CAPM, uses the returns on the market index as proxy to the returns on the market portfolio and defines the return on the stock for the time interval \( t \) to \( t+1 \) as

\[ R_{s,t+1} = (1 - \beta) r_f + \beta R_{M,t+1} + \varepsilon_{s,t+1} \]  

(2.2)

under the assumptions that \( \text{Cov}(\varepsilon_i', R_M) = 0 \) and \( E(\varepsilon_i' \varepsilon_j') = 0, \ i = 1, \ldots, k, \ j = 1, \ldots, k, \ i \neq j \). The disturbance variable, \( \varepsilon_{s,t+1} \), has zero mean and variance equal to \( \sigma_{\varepsilon}^2 \).

We define the total return on the stock for the time interval \( t \) to \( t+1 \) as price returns

\[ R_{S,t+1} = \frac{S_{t+1} - S_t}{S_t}. \]  

(2.3)

We use a market index as proxy to the market portfolio and define the return on the market index for the time interval \( t \) to \( t+1 \) as index returns

\[ R_{M,t+1} = \frac{M_{t+1} - M_t}{M_t}. \]  

(2.4)

\(^1\)This is a simplified version of the CAPM given in Huang and Litzenberger[18, Eqn(10.6.1)], where the return on a portfolio having a zero covariance with respect to the market portfolio is represented by the riskless rate.

\(^2\)This is the version of the Market Model, expressed in terms of Eqn (2.1). See Elton and Gruber[14, p338].

\(^3\)This is referred to as the disturbance term in Huang and Litzenberger[18,p311].
For the time interval $t$ to $t+1$, the disturbance variable is given by

$$\epsilon_{S,t+1}.$$  \hspace{1cm} (2.5)

Using Eqns (2.3) to (2.5), Eqn (2.2) becomes

$$S'_{t+1}(E_{t+1}) = S'_t \left\{ (1 - \beta) r_f + \beta \frac{M_{t+1} - M_t}{M_t} + \epsilon_{t+1} \right\} + S'_t, \quad \text{for } i = 1, \ldots, k \hspace{1cm} (2.6)$$

This is the price determination function that will be used by minimax. Eqn (2.6) is derived directly from the Market Model; we use the name CAPM minimax to indicate that the basic motivation came from the CAPM.

### 2.2 The CAPM-based minimax problem formulation

The problem is to minimize an objective function under a worst case scenario. The minimizing variable is $n$, and the maximizing variable is $E_{t+1}$, which is allowed to take any value within predefined bounds. The minimax problem is

$$\min_{n, E_{t+1}} \max f(n_t, E_{t+1})$$  \hspace{1cm} (2.7)

subject to  \hspace{1cm} $E_l \leq E_{t+1} \leq E_u$ \hspace{1cm} (2.8)

where $f(n_t, E_{t+1})$ is the objective function, discussed in Section 2.3, and $E_l \leq E_{t+1} \leq E_u$ is the range that defines the worst case, discussed in Section 2.4. There are no constraints on $n$, the number of shares to hold at time $t$: non-negative $n$ implies a net holding of shares; negative $n$ implies a net sale of shares.
2.3 The objective function

We define \( U_1 : \mathbb{R}^k \times \mathbb{R}^{k+1} \to \mathbb{R}^1 \), \( U_2 : \mathbb{R}^k \to \mathbb{R}^k \), \( U : \mathbb{R}^k \times \mathbb{R}^{k+1} \to \mathbb{R}^{k+1} \), \( n_i \in \mathbb{R}^k \), \( E_{r+1} \in \mathbb{R}^{k+1} \) and \( Q \) as a \((k+1)\times(k+1)\) positive definite diagonal weighting matrix. \( U_1 \) refers to the potential hedging error for the next period. \( U_2 \) refers to transaction costs associated with buying or selling of stocks. \( U^d \in \mathbb{R}^{k+1} \) is the vector of desired values for the potential hedging error for the two periods and the transaction cost terms: we use a desired value of zero, i.e. the desired hedging error is zero and the desired transaction cost is zero.

We consider the objective function

\[
\begin{aligned}
f(n_1, E_{t+1}) &= \frac{1}{2} < U - U^d, Q(U - U^d) > \\
\text{where} \\
n_i &= \begin{bmatrix} n_i^1 \\ \vdots \\ n_i^k \end{bmatrix} \quad \text{and} \quad E_{t+1} = \begin{bmatrix} M_{t+1}^1 \\ \epsilon_{t+1}^1 \\ \vdots \\ \epsilon_{t+1}^k \end{bmatrix} \\
U(n_1, E_{t+1}) &= \begin{bmatrix} U_1(n_1, E_{t+1}) \\ \cdots \\ U_2(n_1) \end{bmatrix} \quad \text{and} \quad U^d = \begin{bmatrix} U_{t+1}^d \\ \cdots \\ U_{t+2}^d \end{bmatrix} = \begin{bmatrix} 0 \\ \cdots \\ 0 \end{bmatrix}
\end{aligned}
\]

\[
U_1(n_1, E_{t+1}) = \sum_{i=1}^k n_i^t (S_{i+1}^t (E_{t+1}) - S_i^t) + \sum_{i=1}^k N_i^t (B_i^t - B_{i+1}^t (S_{i+1}^t (E_{t+1}))) \\
\quad + \sum_{i=1}^k (- (n_{i-1}^t - n_{i-2}^t) S_{i-1}^t + C_{i-1}^t (1 + r \Delta t)) r \Delta t
\]

where

\[
C_{i-1}^t = C_{i-2}^t (1 + r \Delta t) - (n_{i-1}^t - n_{i-2}^t) S_{i-1}^t - \hat{K}_1 (n_{i-1}^t - n_{i-2}^t) S_{i-1}^t.
\]

\[\text{In minimax, we can adopt any desired value. We adopted a desired value of zero because in delta hedging, the expected value of the hedging error is zero.}\]
\[ U_2(n) = \begin{bmatrix} U_2^1(n_1^i) \\ \vdots \\ U_2^k(n_k^i) \end{bmatrix} \]  
(2.14)

where \( U_2^i(n_i^j) = \tilde{K}(n_i^j - n_{i-1}^j)S_i^j. \)  
(2.15)

We first identify all the variables in Eqns (2.7) to (2.15) and then give the economic interpretation of Eqn (2.12).

\( C_{i-1}^j(1 + r\Delta t) \) is the value of \( C_{i-1}^j \) with interest payments. At time \( t \), \( C_{i-1}^j \) is a constant: all the variables in (2.13) have actual values. We note that \( C_{i-1}^j \) will normally be a negative number.

Transaction costs do not come into the objective function directly as part of \( U_1 \) because they introduce nondifferentiability into the equation. Instead, we introduce \( U_2^j \) to represent a penalty for transaction costs for each option \( i \) at time \( t \).

In the weighting matrix \( Q \), the weights \( q_i, \ i=1,\ldots,k+1 \), which are specified by the hedger, represent his preferences: high \( q_1 \) represents an emphasis on minimizing the potential hedging error. High \( q_i, \ i=2,\ldots,k+1 \), represents an emphasis on minimizing the corresponding transaction cost term.

For each option \( i, i=1,\ldots,k, B_{t+1}^i(S_{t+1}^i(E_{t+1})) \) is valued using the Black and Scholes[1] option pricing model and using Leland's modified volatility estimate[2]. For each stock \( i, i=1,\ldots,k, S_{t+1}^i(E_{t+1}) \) is computed using the price determination function, Eqn (2.6).

\[ ^5 \text{This is discussed in detail in Chapter 4, Section 4.6.} \]
\[ ^6 \text{This is the cumulative value of cash inflow minus cash outflow, first introduced in Chapter 4, Section 4.4.3.1.} \]
\[ ^7 \text{The treatment of transaction costs is discussed in detail Chapter 4, Section 4.6.} \]
\[ ^8 \text{Eqns (4.12), (4.13) and (4.14) in Chapter 4, Section 4.4.2.1.} \]
\[ ^9 \text{Eqn (3.15) in Chapter 4, Section 3.2.} \]
We now give an economic interpretation of (2.12). $U_t$ represents the potential hedging error between time $t$ and time $t+1$: it comprises the potential shift in the stock position, the potential shift in the option position and the potential interest payment. It is a function of the variable $E_{t+1}$ which is a vector of specific error variables and a market variable. We wish to minimize the potential hedging error, including interest payments on borrowed money, using a worst case scenario based on market risk and specific risk.

2.4 The worst case scenario

From (2.10), the source of uncertainty $E_{t+1}$ is defined in terms of $k+1$ variables: the first variable $M_{t+1}$ accounts for the uncertainty due to market movements and the last $k$ variables $e^i_{t+1}$ account for the uncertainty that is specific to stock $i$, $i = 1, ..., k$.

The uncertainty range, as defined by (2.8), can be reformulated as:

\[
M_l \leq M_{t+1} \leq M_u \tag{2.16}
\]

\[
\varepsilon^l_i \leq \varepsilon^i_{t+1} \leq \varepsilon^u_i , \quad i = 1, ..., k \tag{2.17}
\]

In this study we consider two worst case scenarios, given in the following two sections.

2.4.1 Worst Case 1

Worst Case 1 relates to extreme movements in the market index and in specific error variables consistent with the most likely values that they may have within the 95% confidence interval. We define the range of $E_{t+1}$, where we evaluate the effect of the worst case scenario, as the range whose upper and lower bounds delimit the 95% confidence interval of all possible values of the future market index and specific error variables, i.e. two standard deviations about the expected value at time $t+1$. For $M_{t+1}$, the 95% confidence interval will be based on the market risk. For the specific error
variables $\epsilon_{t+i}, \; i=1,...,k$, the 95% confidence interval will be based on the corresponding specific risks.

Worst Case 1 is hereafter referred to as the 95%-Level.

2.4.2 Worst Case 2

Worst Case 2 relates to extreme movements in the market index that may result in a switch in the state of the option from in-the-money to out-of-the-money, or vice versa. A switch in the state of the option may result in a higher hedging error. We define the range of $M_{t+i}$, where we evaluate the effect of the worst case scenario, as the range whose upper and lower bounds delimit the possible values of the future market index within one and three standard deviations from the expected value of the market index at time $t+1$. We choose the side of the distribution that may result in a stock price that is closer to the exercise price $X$. This means that if $S_t > X$, then the relevant range would be on the left side of the distribution of future market index; if $S_t \leq X$, then it would be on the right side.

For a portfolio of stocks, the range of $M_{t+i}$ may be determined by several criteria such as the states of the options (whether in-the-money, at-the-money, or out-of-the-money) and the relative distance of each stock price from the exercise price weighted by the volatility. In this simulation we use the criterion that if the majority of the options are in-the-money, then we use the left side of the distribution of future market index and if the majority of the options are out-of-the-money, then we use the right side. The specific error variables will take their 95% level ranges as defined in Worst Case 1.

Worst Case 2 is hereafter referred to as the Abrupt-Change.$^{10}$

$^{10}$Abrupt-Change refers to the 1 to 3 standard deviation region of the distribution of future levels. For Basic minimax, we consider stock price levels and for CAPM minimax, we consider the market index level.
2.5 The variants of CAPM minimax

We consider two variants of CAPM minimax: the 95%-Level variant, and the Abrupt-Change variant. Both of these worst case scenarios are as described in Section 2.4.

3 Simulation study of the performance of CAPM minimax

In this section, we simulate the relative performance of the two variants of CAPM minimax when they are used to hedge the risk of writing a European call option. We describe the generation of the options and their underlying stock price series in Section 3.2. We selected a set of options for this simulation on the basis of the correlation of the returns on their underlying stock with the returns on a market index. From this set we randomly selected portfolios of 5 options. In our analysis we use the definition of performance and relative performance given in Section 5.6 of Chapter 4.

3.1 The objective of the simulation

This simulation is intended to serve as a feasibility study on potential extensions to Basic minimax. Towards this, the simulation is used to ascertain whether CAPM minimax outperforms Basic minimax for individual options and for portfolios of options.

3.2 Generation of data

We described the method for generating stock price series and option price series in Section 5.2 of Chapter 4. We generated a market index series with which we computed the beta of the stocks. We then created 150 portfolios of 5 options by randomly selecting from all options whose underlying stocks have betas within 0.5 to 1.5. Hereafter, we refer to a simulation run on a portfolio as a replication.
3.3 From set-up to wind-down

In Section 5.4 of Chapter 4, we discussed the mechanics of hedging, from setting up to winding down the hedge. For this simulation, we use the same mechanics with a few changes in parameters. The strategies involve rebalancing the hedge at uniform intervals of time; in the simulation, the interval is one week. Weekly data include the price of the stock, the price of the option, the risk-free interest rate, the time to maturity and the volatility (sigma) of the returns on the stock, given as one of five preset levels: 0.2, 0.3, 0.4, 0.5, and 0.6. The risk-free interest rate is preset at 0.10. Dividends are excluded from the analysis. In the simulation, \( N \), the number of contracted shares, is 100. In addition, for CAPM minimax, the beta of the stocks varied between 0.5 and 1.5.

3.4 Stratification

In this simulation we apply two strategies, Basic minimax and CAPM minimax, to sets of 5 options; first we apply the strategies to the options individually, and then to a portfolio of the 5 options. We also apply delta hedging to the options individually; we did not apply delta hedging to a portfolio of the options because its performance when applied to a portfolio is the summation over its performance when applied to individual options in the portfolio. Table 3.4 summarizes the stratification according to the treatment of options.

<table>
<thead>
<tr>
<th>Name</th>
<th>Strategy</th>
<th>Applied to</th>
<th>Measure of overall performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH</td>
<td>Delta Hedging</td>
<td>Individual options</td>
<td>sum of all the individual performances</td>
</tr>
<tr>
<td>Basic(I)</td>
<td>Basic Minimax</td>
<td>Individual options</td>
<td>sum of all the individual performances</td>
</tr>
<tr>
<td>CAPM(I)</td>
<td>CAPM Minimax</td>
<td>Individual options</td>
<td>sum of all the individual performances</td>
</tr>
<tr>
<td>Basic(P)</td>
<td>Basic Minimax</td>
<td>Portfolio of options</td>
<td>performance for the portfolio</td>
</tr>
<tr>
<td>CAPM(P)</td>
<td>CAPM Minimax</td>
<td>Portfolio of options</td>
<td>performance for the portfolio</td>
</tr>
</tbody>
</table>
Within the minimax strategies, we use two variants that correspond to two worst case scenarios: the 95%-Level and the Abrupt-Change variants. For Basic minimax, these are the same as, respectively, variants E and F in Chapter 4. For CAPM minimax they are as described in Section 2.5.

3.5 Summary of simulation results

In this section, we present the results of the simulation. In Table 3.5a, we summarize the relative performance of the strategies applied to different treatments of options when the worst case scenario is defined by the 95%-Level and by Abrupt-Change. In Table 3.5b, we summarize the difference in performance between minimax strategies. The tables also contain, in italics in small font, the absolute value of the t-statistics, followed by ** if the difference is significant at the 2% level, and * if it is significant at the 10% level.

Table 3.5a Relative performance of strategies (t values in italics)

Units: Percentage points of final cumulative value of cash inflow minus cash outflow divided by the notional value of the contract

<table>
<thead>
<tr>
<th>STRATEGY</th>
<th>Worst case scenario</th>
<th>Basic(I)</th>
<th>Basic(P)</th>
<th>CAPM(I)</th>
<th>CAPM(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%-Level</td>
<td>1.9, 0.8, 1.7, 2.1</td>
<td>2.1, 1.9, 1.8, 2.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abrupt-Change</td>
<td>2.1, 1.9, 1.8, 2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.5b Difference in performance between minimax strategies.(t values in italics)

Units: Percentage points of final cumulative value of cash inflow minus cash outflow divided by the notional value of the contract

<table>
<thead>
<tr>
<th>STRATEGY</th>
<th>Worst case scenario</th>
<th>CAPM(P) minus Basic(P)</th>
<th>CAPM(P) minus CAPM(I)</th>
<th>CAPM(P) minus Basic(I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%-Level</td>
<td>1.3, 0.4, 0.2, -0.2</td>
<td>0.9, 0.0, -1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abrupt-Change</td>
<td>0.2, 0.3, 0.0, -0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We use significance levels of 2% and 10% to express the results of a 2-tailed test; however, we use a 1-tailed test for the sign of the difference at the 5% level.

11We use significance levels of 2% and 10% to express the results of a 2-tailed test; however, we use a 1-tailed test for the sign of the difference at the 5% level.
3.5.1 Main results from Table 3.5a and Table 3.5b

In this section we give the main results from Table 3.5a and Table 3.5b.

From Table 3.5a, for the Abrupt-Change scenario, the relative performance of the strategies is much the same; each strategy outperforms DH by about 2 percentage points, with CAPM(P) being joined by Basic(I) as the best performers. For the 95%-Level scenario, although all strategies outperform DH, there is a greater variation in relative performance, with CAPM(P) being the best performer, and with Basic(P) being the worst performer.

From Table 5.3b, the biggest differences in performance occurred under the 95%-Level scenario, with the biggest positive difference occurring for CAPM(P) minus Basic (P), and the largest negative difference occurring for Basic(P) minus Basic(I).

3.6 Statistical testing of hypotheses

In this section we formulate and test hypotheses; we also comment on any rejected hypotheses. We test a number of hypotheses on the difference in relative performance of the minimax strategies. In Table 3.6, we list 10 hypotheses, and give their accept/reject state at the 5% level of significance\(^{12}\).

Because all the strategies considered in this section are based on two of the high performing variants of Basic minimax (E and F), we hypothesize (H1, H2, H3, and H4) that each strategy outperforms DH.

Because the worst case scenarios of CAPM(I) reduce to the worst case scenarios of Basic(I), we hypothesize (H5) that their performances are not significantly different.

\(^{12}\)We use the 5% level of significance to reflect a less stringent criterion on the acceptance or rejection of a hypothesis because the simulation data exhibit a wide range of moneyness and the hypotheses we tested are not sufficiently refined to examine the effect of variations in moneyness.
Because the portfolio's worst case scenario may be the result of a compounding of the worst case scenarios for the individual options, we hypothesize (H6) that Basic(I) outperforms Basic(P).

Because a portfolio consisting of the underlying stocks that are positively correlated with the market index may, at the same time, have low specific risk because of diversification, the worst case effect of the specific error variables may be diversified away. This implies that the worst case scenario for the portfolio may be less extreme than the worst case scenarios for the individual options. We therefore hypothesize (H7) that CAPM(P) outperforms CAPM(I). For the same reason, we hypothesize (H8) that CAPM(P) outperforms Basic(I), and by transitivity (from H6 and H8), we hypothesize (H9) that CAPM(P) outperforms Basic(P).

Because the portfolio's worst case scenario under Basic minimax may be the result of a compounding of the worst cases for individual options, we hypothesize (H10) that CAPM(I) outperforms Basic(P).

Table 3.6 Accept/reject states of the hypotheses\(^\text{13}\). (R = Reject)

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>95%-Level</th>
<th>Abrupt-Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Basic(I) outperforms DH</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H2: Basic(P) outperforms DH</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H3: CAPM(I) outperforms DH</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H4: CAPM(P) outperforms DH</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H5: Basic(I) is not different from CAPM(I)</td>
<td>Accept</td>
<td>R</td>
</tr>
<tr>
<td>H6: Basic(I) outperforms Basic(P)</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H7: CAPM(P) outperforms CAPM(I)</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H8: CAPM(P) outperforms Basic(I)</td>
<td>Accept</td>
<td>R</td>
</tr>
<tr>
<td>H9: CAPM(P) outperforms Basic(P)</td>
<td>Accept</td>
<td>Accept</td>
</tr>
<tr>
<td>H10: CAPM(I) outperforms Basic(P)</td>
<td>Accept</td>
<td>R</td>
</tr>
</tbody>
</table>

\(^{13}\)Except for Hypothesis 5, we tested the hypotheses by defining the Null Hypothesis as no significant difference between means; the Alternative Hypothesis corresponds to the hypothesis outlined above. In Table 2, an accept state means an acceptance of the Alternative Hypothesis.
3.6.1 Comments on the rejected hypotheses

H5 was rejected, with Basic(I) outperforming CAPM(I). A possible reason for this may be a difference in the size of the ranges, with CAPM minimax generating wider ranges that resulted in a worse hedging error than Basic minimax.

The rejection of H8 and H10 under the Abrupt-Change worst case scenario may be explained by the fact the Abrupt-Change is already an extreme case; this may imply that any possible variation in performance when using either of the strategies is insignificant compared to the extreme case condition of the Abrupt-Change.

3.7 Rank ordering

The simulation results suggest the following rank order of positive differences in performance. Table 3.7a gives the rank order for an individual option, and Table 3.7b gives the rank order for a portfolio. These tables are based on Table 3.5a and Table 3.5b.

Table 3.7a  Rank order of positive significant differences in performance for an individual option.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Worst case scenario</th>
<th>95% Level</th>
<th>Abrupt-change</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic (Individual)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CAPM (Individual)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Delta Hedging</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 3.7b  Rank order of positive significant differences in performance for a portfolio of options.

<table>
<thead>
<tr>
<th>Worst case scenario</th>
<th>95% Level</th>
<th>Abrupt-change</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPM (Portfolio)</td>
<td>1</td>
<td>1.5</td>
<td>1.25</td>
</tr>
<tr>
<td>Basic (Individual)</td>
<td>2</td>
<td>1.5</td>
<td>1.75</td>
</tr>
<tr>
<td>CAPM (Individual)</td>
<td>3</td>
<td>3.5</td>
<td>3.25</td>
</tr>
<tr>
<td>Basic (Portfolio)</td>
<td>4</td>
<td>3.5</td>
<td>3.75</td>
</tr>
<tr>
<td>Delta Hedging</td>
<td>5</td>
<td>5</td>
<td>5.00</td>
</tr>
</tbody>
</table>

For individual options and for portfolios, all minimax strategies for the two worst case scenarios outperform delta hedging. For individual options, Basic(I) outperforms CAPM(I). In contrast, for portfolios, CAPM(P) outperforms Basic(I) for the 95%-Level scenario; however, the difference between them is not significant at the 1% level. This difference suggests that CAPM(P) is sensitive to market index levels and may be the most suitable strategy for portfolios. Whereas CAPM(P) outperforms CAPM(I), Basic(I) outperforms Basic(P), especially under the 95% level scenario. These results suggest that CAPM(P) has succeeded in dealing with the problem of the compounding of worst case scenarios that occurs under Basic(P). However, our results are not adequate to establish that CAPM(P) is a superior strategy to Basic(I).

4 The beta of the hedge portfolio for CAPM minimax

The beta of a stock is a coefficient that relates market returns and stock returns. This relationship is given by Eqn (2.4). The true beta of a stock can be estimated using regression analysis. The beta of an option14 is a coefficient that relates market returns and option returns. Because the value of an option is dependent on the value of the stock, the beta of the option is expressed in terms of the stock price.

\[ \beta_B = \frac{\Theta(d_1) * S * \beta_S}{B} \]  

(4.1)

where \( \beta_B \)  beta of the option  
\( \beta_S \)  beta of the stock.

This equation shows that the returns on the option are affected by the returns on the market via the stock and the stock beta. The **beta of the hedge portfolio** is the sum of the betas of its components (stocks and options) weighted by the proportion of the total value of the hedge portfolio contributed by the individual components.

We illustrate the performance of CAPM minimax for a randomly selected replication from the simulation. We present a graph showing the variation with time of the beta of the hedge portfolio. In Figure 4, the graph shows a horizontal line that is the hedge portfolio's beta-risk profile under delta hedging, and an upper line graph that is the market index level. The figure also shows that, unlike delta hedging, CAPM minimax gives a nonzero-beta hedge portfolio, and that it resulted in an increasing beta when the market was falling and in a decreasing beta when the market was rising, with the highest beta occurring at the time when the market is at its lowest (see arrow). This pattern illustrates the performance of the strategy over time: in a falling market, CAPM minimax gives a hedge recommendation in anticipation of a rise in the market index, and in a rising market, in anticipation of a fall in the market index.

**Figure 4** Beta-risk profile using CAPM minimax.
5 Conclusion

We have presented an extension of the basic minimax hedging strategy that uses information about the movements in the market index in the calculation of the potential hedging error; we have also simulated its performance. The results of the simulation suggest that CAPM(P) is sensitive to movements in the market index, and has succeeded in dealing with a major difficulty in the Basic(P) strategy (described in Section 3), a difficulty that was most conspicuous under the 95% level worst case scenario.

The variants of CAPM minimax have been shown to perform well in the circumstances they have been designed for. The results of the simulation suggest that CAPM minimax may be suitable for other hedging problems where the securities are highly sensitive to market movements.
Performance of the Algorithm as a Computational Tool for the Hedging Strategies

1 Introduction

An important distinction is made in Chapter 4 between the performance of the minimax algorithm and the performance of the minimax hedging strategy. In this chapter, we present the performance of the minimax algorithm, applied to risk management in finance, from the point of view of optimization. Chapters 4, 5 and 6 deal with the performance of the minimax hedging strategy from the point of view of risk management.

We briefly introduce the hedging problem and the scenario under which the minimax algorithm is to be used. The scenario would be a description of the risk management policy and the financial variables at a rebalancing date; the rebalancing date is the time that the minimax algorithm is applied in order to implement the risk management policy.

In Section 2, we use the basic minimax strategy as a risk management policy. At a particular rebalancing date, we use the minimax algorithm in order to implement this policy. We present the objective function and its Hessians\(^1\), and relate the Hessians to the type of solution that we can expect algorithm to find, i.e. whether the solution is an extreme point or a mid-range solution. We give three illustrations to demonstrate the performance of the algorithm under the basic minimax strategy.

\(^{1}\text{This is necessary to show the components of the Hessian that account for convexity and concavity.}\)
In Section 3, we use Two-Period minimax as a risk management policy. As in Section 2, we present the objective function and its Hessians. We give four illustrations to demonstrate the performance of the algorithm under Two-Period minimax.

In Section 4, we use CAPM minimax as a risk management policy. As in Section 2, we present the objective function and its Hessians. We present one illustration to demonstrate the performance of the algorithm under CAPM minimax.

In Section 5, we discuss the performance of the algorithm when it is applied to a portfolio of options as compared to when it is applied to individual options and show the superiority of the minimax solution for the portfolio case over the minimax solutions for the individual cases.

2 The performance of the algorithm when used in the basic minimax strategy

In this section, we present the basic minimax strategy as a risk management policy and take a particular rebalancing date to illustrate the application of the algorithm in order to implement this policy. This strategy defines the worst case scenario in terms of extreme movements in the price of the underlying stock. Details of the basic minimax hedging strategy are given in Chapter 4. The minimax formulation and the objective function are presented here for ease of reference.

The minimax problem:

\[
\min_{n_t} \max_S f(n_t, S) \quad (2.1)
\]

subject to \( S_{t+1} \leq S_{t+1} \leq S_{t+1} \) \quad (2.2)

2.1 Objective Function 2 of the basic minimax strategy

We define \( U_1 : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^1 \), \( U_2 : \mathbb{R}^k \rightarrow \mathbb{R}^k \), \( U : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^{k+1} \), \( n_t \in \mathbb{R}^k \), \( S_{t+1} \in \mathbb{R}^k \) and \( Q \) as a \((k + 1) \times (k + 1)\) positive definite diagonal weighting matrix. \( U^d \in \mathbb{R}^{k+1} \) is the
vector of desired values for the potential hedging error and the transaction cost terms: we use a desired value of zero, i.e. the desired hedging error is zero and the desired transaction cost is zero.

\[ f(n_t, S_{t+1}) = \frac{1}{2} <U - U^d, Q(U - U^d)> \]  

(2.3)

where

\[ n_t = \begin{bmatrix} n_t^1 \\ \vdots \\ n_t^k \end{bmatrix} \quad \text{and} \quad S_{t+1} = \begin{bmatrix} S_{t+1}^1 \\ \vdots \\ S_{t+1}^k \end{bmatrix} \]  

(2.4)

\[ U(n_t, S_{t+1}) = \begin{bmatrix} U_1(n_t, S_{t+1}) \\ \vdots \\ U_k(n_t) \end{bmatrix} \quad \text{and} \quad U^d = \begin{bmatrix} U^d_1 \\ \vdots \\ U^d_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \]  

(2.5)

\[ U_1(n_t, S_{t+1}) = \sum_{i=1}^k n_t^i (S_{t+1}^i - S_t^i) + \sum_{i=1}^k N^i (B^i_t - B^i_{t+1}(S^i_{t+1})) + \sum_{i=1}^k (n_t^i - n_{t-1}^i) S_t^i + C_{t-1}^i (1 + r\Delta t) r\Delta t \]  

(2.6)

\[ U_2(n_t) = \begin{bmatrix} U_2^1(n_t^1) \\ \vdots \\ U_2^k(n_t^k) \end{bmatrix} \]  

(2.7)

where

\[ U_2^i(n_t^i) = \tilde{K}(n_t^i - n_{t-1}^i) S_t^i \]  

(2.8)

2.2 The Hessians

In this section, we distinguish between three Hessians: \textit{Hessian}_n which is the Hessian with respect to \( n_t \), \textit{Hessian}_S which is the Hessian with respect to \( S_{t+1} \); and \textit{Hessian}_{BFGS} which is the BFGS approximation to the Hessian of the max-function. \textit{Hessian}_S gives us full information about the curvature of \( f(n_t, S_{t+1}) \) in the \( S_{t+1} \)-space whereas \textit{Hessian}_n does not give full information about the curvature of \( \Phi(n_t) \) in the \( n_t \)-
space. Instead, Hessian_{nn} gives partial information related to the curvature of \( f(n_t, S_{t+1}) \) in the \( n_t \)-space. Evaluation of the Hessian of \( \Phi(n_t) \) requires complete evaluation, and linear combination, of all Hessians for all maximizers. We note that this is very difficult, if not impossible, to calculate due to the need to identify all maximizers and the optimal coefficients for the required combination of Hessians. We therefore present the numerically evaluated Hessian_{nn} as a partial proof towards realizing the curvature of \( \Phi(n_t) \) in the \( n_t \)-space. In the implementation of the quasi-Newton Algorithm, we used the BFGS Hessian approximation to that of \( \Phi(n_t) \) in the \( n_t \)-space.

**Hessian_{nn}:** The objective function is quadratic and the resulting Hessian with respect to \( n_t \) is always positive definite. This implies that the objective function is convex in the \( n_t \)-space.

For \( i = 1, \ldots, k \):
\[
\frac{\partial^2 f}{\partial (n'_t)^2} = q_1(S'_{t+1} - S'_t(1 + r\Delta t))^2 + q_{t+1}(KS'_t)^2
\]  
(2.9)

For \( i = 1, \ldots, k, j = 1, \ldots, k, i \neq j \):
\[
\frac{\partial^2 f}{\partial (n'_t)(n'_t)} = q_1(S'_{t+1} - S'_t(1 + r\Delta t))(S'_{i+1} - S'_i(1 + r\Delta t))
\]  
(2.10)

**Hessian_{SS}:** The Hessian with respect to \( S_{t+1} \) is dependent on the second derivative of the option pricing model.

For \( i = 1, \ldots, k \):
\[
\frac{\partial^2 f}{\partial (S'_{t+1})^2} = \left(-N'_i \frac{\partial B'_{i+1}}{\partial S'_{t+1}} + n'_i \right)^2 + U_i \left(-N'_i \frac{\partial^2 (B'_{i+1})}{\partial (S'_{t+1})^2} \right)
\]  
(2.11)

For \( i = 1, \ldots, k, j = 1, \ldots, k, i \neq j \):
\[
\frac{\partial^2 f}{\partial (S'_{t+1})(S'_i)} = \left(-N'_i \frac{\partial B'_{i+1}}{\partial S'_{t+1}} + n'_i \right) \left(-N'_i \frac{\partial B'_{i+1}}{\partial S'_i} + n'_i \right)
\]  
(2.12)

where
\[
\frac{\partial^2 \left( \Theta_{d, \sigma} \right)}{\partial (S, \sigma)^2} = \frac{\Theta(d_i)}{S_{t+1} \sigma \sqrt{T-t-1}}
\]  
(2.13)
is the second derivative of the option pricing model with

\[ \Theta'(d_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{d_1^2}{2}} \]  

and

\[ d_i = \frac{\ln(S_i/S_0) + (r + \frac{\sigma^2}{2})(T-t-1)}{\sigma\sqrt{T-t-1}} \]  

Eqn (2.13) is always positive. However, it appears as a negative component of the elements of the Hessian. The diagonal elements (2.11) may become negative when the value of \( U_1 \) is positive. In this situation, some of the eigenvalues of the Hessian would be negative and the solution may have a mid-range element.

2.3 Illustrations

The illustrations show the same output variables as outlined in Chapter 3 including the Hessians of the objective function at the solution and the minimax hedging error \( U_i(n_{ret},S_{ret}) \). The first illustration shows the more likely situation when the minimax algorithm is applied in the hedging strategy. When the minimax hedging error \( U_i(n_{ret},S_{ret}) \) is negative, which is the more likely event, the solution is an extreme point. The second illustration shows a case where the minimax hedging error \( U_i(n_{ret},S_{ret}) \) is positive; in this case, the solution is in the middle of the range. The last illustration shows the performance of the minimax algorithm when applied to a portfolio of 5 options.

In the illustrations, we use data from the empirical study in Chapter 4. The rebalancing date is December 4, 1991. The expiration date for all five options is August 1992. Their exercise prices, as well as other parameters, are given in Section 6 of Chapter 4.
Illustration 2.1: Extreme point solution

Hedging the risk of writing an option on the British Telecom Stock:

\[ n_t \in \mathbb{R}^1, \quad S \in \mathbb{R}^1 \]

Objective function:

\[ f(n_t, S) \]

subject to

\[ 3.289920e + 02 \leq S_{t+1} \leq 3.686350e + 02 \]

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwiell’s Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
</tbody>
</table>

\[ n_t = 6.185360 e+01 \]

\[ S_{t+1} = 3.470000 e+02 \]

\[ ldl = 1.760497 e-16 \]

\[ \Psi = 0.000000 e+00 \]

\[ \Phi(n_{t*}) = 1.508524 e+04 \]

\[ n_{t*} = 6.303667 e+01 \]

\[ S_{t+1*} = 3.289920 e+02 \]

\[ k_{\alpha} l \alpha_k = 1; \forall k > k_{\alpha} \]

No. of iterations: 5

Time: 0.3 sec

Eigenvalues of numerically evaluated
\[ \text{Hessian}_n \]

5.166124 e+03

3.111503 e+03

2.583062 e+03

Eigenvalues of numerically evaluated
\[ \text{Hessian}_S \]

5.166124 e+03

3.111503 e+03

2.583062 e+03

Eigenvalues of Hessian approximation
\[ \text{Hessian}_{BFGS} \]

5.166124 e+03

3.111503 e+03

2.583062 e+03

Minimax Hedging Error: -168 points

Time: 0.3 sec
Illustration 2.2: mid-range solution

Hedging the risk of writing an option on the Cadbury Stock:

\[ n_t \in \mathbb{R}^1, \quad S \in \mathbb{R}^1 \]

Objective function:

\[ f(n_t, S) \]

subject to

\[ 4.354752e + 02 \leq S_{t+1} \leq 4.831153e + 02 \]

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwiels's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>Epsilon ( \epsilon )</td>
<td>1.000000 e-06</td>
<td>1.000000 e-06</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
<tr>
<td>( n_t )</td>
<td>6.554278 e+01</td>
<td>6.554278 e+01</td>
</tr>
<tr>
<td>( S_{t+1} )</td>
<td>4.570000 e+02</td>
<td>4.570000 e+02</td>
</tr>
<tr>
<td>( ldl )</td>
<td>2.591688 e-04</td>
<td>2.591688 e-04</td>
</tr>
<tr>
<td>( \Psi )</td>
<td>-1.584881 e-08</td>
<td>-1.584881 e-08</td>
</tr>
<tr>
<td>( \Phi(n_{t+1}) )</td>
<td>1.053631 e+05</td>
<td>1.053631 e+05</td>
</tr>
<tr>
<td>( n_{t+1} )</td>
<td>6.763951 e+01</td>
<td>6.763951 e+01</td>
</tr>
<tr>
<td>( S_{t+1} )</td>
<td>4.602877 e+02</td>
<td>4.602877 e+02</td>
</tr>
<tr>
<td>( k_{\alpha} )</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>No. of iterations</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Time</td>
<td>0.1 sec</td>
<td>0.1 sec</td>
</tr>
<tr>
<td>Eigenvalues of numerically evaluated Hessian ( nn )</td>
<td>8.359353 e+03</td>
<td>8.359353 e+03</td>
</tr>
<tr>
<td>Eigenvalues of numerically evaluated Hessian ( ss )</td>
<td>-2.954345 e+02</td>
<td>-2.954345 e+02</td>
</tr>
<tr>
<td>Eigenvalues of Hessian approximation ( BFGS )</td>
<td>4.535665 e+03</td>
<td>4.535665 e+03</td>
</tr>
<tr>
<td>Minimax Hedging Error</td>
<td>459 points</td>
<td>459 points</td>
</tr>
</tbody>
</table>
Illustration 2.3: extreme point solution

Hedging the risk of writing 5 options:
Guinness, Prudential, British Telecom, Thames Water and Tesco:

\[ n_t \in \mathbb{R}^5 \quad S \in \mathbb{R}^5 \]

Objective function:
\[ f(n_t, S) \]

subject to

\[ 4.731038e+02 \leq S^1_{t+1} \leq 5.301120e+02 \]
\[ 2.142715e+02 \leq S^2_{t+1} \leq 2.400908e+02 \]
\[ 3.289920e+02 \leq S^3_{t+1} \leq 3.686350e+02 \]
\[ 3.185629e+02 \leq S^4_{t+1} \leq 3.569492e+02 \]
\[ 2.247006e+02 \leq S^5_{t+1} \leq 2.517767e+02 \]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-06</td>
<td>1.000000 e-06</td>
<td>1.000000 e-06</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td></td>
</tr>
</tbody>
</table>

\[ n_t \]
\[ 4.807649 \ e+01 \quad 4.807649 \ e+01 \quad 4.807649 \ e+01 \]
\[ 7.484878 \ e+01 \quad 7.484878 \ e+01 \quad 7.484878 \ e+01 \]
\[ 6.185360 \ e+01 \quad 6.185360 \ e+01 \quad 6.185360 \ e+01 \]
\[ 7.318315 \ e+01 \quad 7.318315 \ e+01 \quad 7.318315 \ e+01 \]
\[ 8.275244 \ e+01 \quad 8.275244 \ e+01 \quad 8.275244 \ e+01 \]

\[ S_{t+1} \]
\[ 4.990000 \ e+02 \quad 4.990000 \ e+02 \quad 4.990000 \ e+02 \]
\[ 2.260000 \ e+02 \quad 2.260000 \ e+02 \quad 2.260000 \ e+02 \]
\[ 3.470000 \ e+02 \quad 3.470000 \ e+02 \quad 3.470000 \ e+02 \]
\[ 3.360000 \ e+02 \quad 3.360000 \ e+02 \quad 3.360000 \ e+02 \]
\[ 2.370000 \ e+02 \quad 2.370000 \ e+02 \quad 2.370000 \ e+02 \]

\[ \mid dl \]
\[ 1.245695 \ e-04 \quad 1.245695 \ e-04 \quad 1.416248 \ e-02 \]
\[ \Psi \]
\[ -7.607778 \ e-07 \quad -7.607778 \ e-07 \quad -2.005757 \ e-04 \]
\[ \Phi(n_{t+}) \]
\[ 4.747561 \ e+05 \quad 4.747561 \ e+05 \quad 4.910554 \ e+05 \]

\[ n_{t+} \]
\[ 4.723223 \ e+01 \quad 4.723223 \ e+01 \quad 4.725500 \ e+01 \]
\[ 7.593698 \ e+01 \quad 7.593698 \ e+01 \quad 7.536454 \ e+01 \]
\[ 6.266327 \ e+01 \quad 6.266327 \ e+01 \quad 6.266036 \ e+01 \]
\[ 7.369503 \ e+01 \quad 7.369503 \ e+01 \quad 7.369543 \ e+01 \]
\[ 7.657575 \ e+01 \quad 7.657575 \ e+01 \quad 8.004109 \ e+01 \]

211
<table>
<thead>
<tr>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>$S_{r+1}$ *</td>
<td>5.301120 e+02</td>
</tr>
<tr>
<td></td>
<td>2.142715 e+02</td>
</tr>
<tr>
<td></td>
<td>3.289920 e+02</td>
</tr>
<tr>
<td></td>
<td>3.569492 e+02</td>
</tr>
<tr>
<td></td>
<td>2.517767 e+02</td>
</tr>
</tbody>
</table>

$k_0 | \alpha_k = 1; \forall k > k_0$

| | 0 | 0 | 0 |
| No. of iterations | 36 | 36 | 749 |
| Time | 4.3 sec | 4.3 sec | 115.6 sec |
| Remark | FAILURE: alpha=0 |

Eigenvalues of numerically evaluated $Hessian_{nn}$

| | 2.096500 e+03 | 2.096500 e+03 | 2.095970 e+03 |
| | 2.435039 e+03 | 2.435039 e+03 | 2.429791 e+03 |
| | 4.648141 e+03 | 4.648141 e+03 | 4.635121 e+03 |
| | 5.384713 e+03 | 5.384713 e+03 | 5.474271 e+03 |
| | 1.103836 e+04 | 1.103836 e+04 | 1.105636 e+04 |

Eigenvalues of numerically evaluated $Hessian_{SS}$

| | 3.552895 e+02 | 3.552895 e+02 | 3.306053 e+02 |
| | 3.639348 e+02 | 3.639348 e+02 | 3.395314 e+02 |
| | 5.165098 e+02 | 5.165098 e+02 | 3.979137 e+02 |
| | 7.825777 e+02 | 7.825777 e+02 | 6.642355 e+02 |
| | 1.394759 e+03 | 1.394759 e+03 | 1.202099 e+03 |

Eigenvalues of Hessian approximation $Hessian_{BFGS}$

| | 5.602258 e+03 | 5.602258 e+03 |
| | 2.193895 e+04 | 2.193895 e+04 |
| | 9.472012 e+04 | 9.472012 e+04 |
| | 4.696796 e+06 | 4.696796 e+06 |
| | 3.849028 e+07 | 3.849028 e+07 |

Minimax Hedging Error

| -876 points | -876 points | -818 points |
3 The performance of the algorithm when used in Two-Period minimax

In this section, we present Two-Period minimax as a risk management policy and take a particular rebalancing date to illustrate the application of the algorithm in order to implement this policy. This strategy defines a worst case scenario for two time periods. The difference between this two-period worst case and the one-period worst case is that the range for the future stock price $S_{t+2}$ is wider for the two-period case. This would result in a higher worst-case potential hedging error and a more cautious hedge recommendation. Details of Two-Period minimax are given in Chapter 5. The minimax formulation and the objective function are presented here for ease of reference.

The minimax problem:

\[
\min_{n_t} \max_S f(n_t, S) \quad (3.1)
\]

subject to

\[
S_{t+1} \leq S_{t+1} \leq S_{t+1} \quad (3.2)
\]

\[
S_{t+2} \leq S_{t+2} \leq S_{t+2} \quad (3.2)
\]

3.1 The objective function of Two-Period minimax

We define $U_1: \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^1$, $U_2: \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^1$, $U_3: \mathbb{R}^k \rightarrow \mathbb{R}^k$, $U: \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^{k+2}$, $n_t \in \mathbb{R}^k$, $S_{t+1} \in \mathbb{R}^k$, $S_{t+2} \in \mathbb{R}^k$ and $Q$ as a $(k+2)(k+2)$ positive definite diagonal weighting matrix. $U_1$ refers to the potential hedging error for the first time period and $U_2$ refers to that for the second time period. $U_3$ refers to a penalty term for transaction costs associated with buying or selling of stocks. $U^d \in \mathbb{R}^{k+2}$ is the vector of desired values for the potential hedging error for the two periods and the transaction cost terms: we use a desired value of zero, i.e. the desired hedging error is zero and the desired transaction cost is zero.

The objective function is given by

\[
f(n_t, S_{t+1}, S_{t+2}) = \frac{1}{2} < U - U^d, Q(U - U^d) > \quad (3.3)
\]
where

\[ n_i = \begin{bmatrix} n_i^1 \\ \vdots \\ n_i^k \end{bmatrix} \quad S_{i+1} = \begin{bmatrix} S_{i+1}^1 \\ \vdots \\ S_{i+1}^k \end{bmatrix} \quad S_{i+2} = \begin{bmatrix} S_{i+2}^1 \\ \vdots \\ S_{i+2}^k \end{bmatrix} \] (3.4)

\[ U(n_i, S_{i+1}, S_{i+2}) = \begin{bmatrix} U_1(n_i, S_{i+1}) \\ U_2(n_i, S_{i+2}) \\ U_3(n_i) \end{bmatrix} \quad \text{and} \quad U^d = \begin{bmatrix} U_1^d \\ U_2^d \\ U_3^d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \] (3.5)

\[ U_1(n_i, S_{i+1}) = \sum_{i=1}^k n_i^i (S_{i+1}^i - S_i^i) + \sum_{i=1}^k N^i (B_i^i - B_{i+1}(S_{i+1}^i)) + \sum_{i=1}^k (-n_i^i - n_{i-1}^i) S_i^i + C_{i-1} (1 + r \Delta t) r \Delta t \] (3.6)

\[ U_2(n_i, S_{i+1}, S_{i+2}) = \sum_{i=1}^k n_i^i (S_{i+2}^i - S_{i+1}^i) + \sum_{i=1}^k N^i (B_i^i - B_{i+1}(S_{i+1}^i) - B_{i+2}(S_{i+2}^i)) \]

\[ + \sum_{i=1}^k (-n_i^i - n_{i-1}^i) S_i^i + C_{i-1} (1 + r \Delta t)(1 + r \Delta t)(r \Delta t) \] (3.7)

\[ U_3(n_i) = \begin{bmatrix} U_1^i(n_i^1) \\ \vdots \\ U_3^i(n_i^k) \end{bmatrix} \] (3.8)

where

\[ U_3^i(n_i) = \hat{K}(n_i^i - n_{i-1}^i) S_i^i \] (3.9)

### 3.2 The Hessians

In this section, we distinguish between three Hessians: Hessian\(_{nn}\), which is the Hessian with respect to \(n_i\); Hessian\(_{SS}\), which is the Hessian with respect to \(S_{i+1}\); and Hessian\(_{BFGS}\), which is the BFGS approximation to the Hessian of the max-function. The remarks in Section 2.2 apply.
\textbf{Hessian}_{nn}: The objective function is quadratic and the resulting Hessian with respect to \(n_t\) is always positive definite. This implies that the objective function is convex in the \(n_t\)-space.

For \(i = 1, \ldots, k\):
\[
\frac{\partial^2 f}{\partial (n_t^i)^2} = q_1 (S_{t+1}^i - S_t^i (1 + r \Delta t))^2 + q_2 ((S_{t+2}^i - S_{t+1}^i - S_t^i (1 + r \Delta t)) r \Delta t)^2 + q_{1/2} (K S_t^i)^2
\]  
(3.10)

For \(i = 1, \ldots, k\), \(j = 1, \ldots, k\), \(i \neq j\):
\[
\frac{\partial^2 f}{\partial (n_t^i)(n_t^j)} = q_1 (S_{t+1}^i - S_t^i (1 + r \Delta t))(S_{t+1}^j - S_t^j (1 + r \Delta t)) + q_2 ((S_{t+2}^i - S_{t+1}^i - S_t^i (1 + r \Delta t)) r \Delta t)((S_{t+2}^j - S_{t+1}^j - S_t^j (1 + r \Delta t)) r \Delta t)
\]  
(3.11)

\textbf{Hessian}_{SS}: The Hessian with respect to \(S_{t+1}^i\) is dependent on the second derivative of the option pricing model.

For \(i = 1, \ldots, k\):
\[
\frac{\partial^2 f}{\partial (S_{t+1}^i)^2} = \left( -\frac{\partial B_{t+1}^i}{\partial S_{t+1}^i} + n_t^i \right)^2 + U_1 \left( \frac{\partial^2 (B_{t+1}^i)}{\partial (S_{t+1}^i)^2} + \frac{\partial B_{t+1}^i}{\partial S_{t+1}^i} - n_t^i \right)^2 + U_2 \left( \frac{\partial^2 (B_{t+1}^i)}{\partial (S_{t+1}^i)^2} \right)
\]  
(3.12)

For \(i = 1, \ldots, k\), \(j = k + 1, \ldots, 2k\):
\[
\frac{\partial^2 f}{\partial (S_{t+1}^i)(S_{t+2}^j)} = \left( -\frac{\partial B_{t+1}^i}{\partial S_{t+1}^i} + n_t^i \right) \left( -\frac{\partial B_{t+2}^j}{\partial S_{t+2}^j} + n_t^j \right) + U_1 \left( \frac{\partial B_{t+1}^i}{\partial S_{t+1}^i} - n_t^i \right) \left( \frac{\partial B_{t+2}^j}{\partial S_{t+2}^j} - n_t^j \right)
\]  
(3.13)

For \(i = k + 1, \ldots, 2k\):
\[
\frac{\partial^2 f}{\partial (S_{t+2}^i)^2} = \left( -\frac{\partial B_{t+2}^i}{\partial S_{t+2}^i} + n_t^i \right)^2 + U_2 \left( \frac{\partial^2 (B_{t+2}^i)}{\partial (S_{t+2}^i)^2} \right)
\]  
(3.14)

For \(i = 1, \ldots, k\), \(j = 1, \ldots, k\), \(i \neq j\):
\[
\frac{\partial^2 f}{\partial (S_{t+1}^i)(S_{t+1}^j)} = \left( \frac{\partial B_{t+1}^i}{\partial S_{t+1}^i} + n_t^i \right) \left( \frac{\partial B_{t+1}^j}{\partial S_{t+1}^j} + n_t^j \right)
\]  
(3.15)
For \(i = k + 1, \ldots, 2k\), \(j = k + 1, \ldots, 2k\), \(i \neq j\)

\[
\frac{\partial^2 f}{\partial (S'_{i}) (S'_{j})} = \left( \frac{\partial B'_{i}}{\partial S'_{i}} - n'_{i} \right) \left( \frac{\partial B'_{j}}{\partial S'_{j}} + n'_{j} \right)
\]  (3.16)

For \(j = k + 1, \ldots, 2k\), \(i = k + 1, \ldots, 2k\), \(i \neq j\)

\[
\frac{\partial^2 f}{\partial (S'_{j}) (S'_{i})} = \left( \frac{\partial B'_{j}}{\partial S'_{j}} - n'_{j} \right) \left( \frac{\partial B'_{i}}{\partial S'_{i}} + n'_{i} \right)
\]  (3.17)

where
\[
\frac{\partial^2 (\Theta)}{\partial (S', S'_{j})^2}
\]  (3.18)

is calculated using Eqns (2.13) to (2.15), and

\[
\frac{\partial^2 (\Theta)}{\partial (S', S'_{j})^2} = \frac{\Theta'(d_{j})}{S'_{j} \sigma \sqrt{T - t - 2}}
\]  (3.19)

is the second derivative of the option pricing model with

\[
\Theta'(d_{j}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{d_{j}^2}{2}}
\]  (3.20)

and

\[
d_{j} = \frac{\ln(S'_{j}/x) + (r + \frac{\sigma^2}{2})(T - t - 2)}{\sigma \sqrt{T - t - 2}}
\]  (3.21)

Eqn (3.18) is always positive. However, it appears in (3.12) as a negative component of the second term and as a positive component of the fourth term. Eqn (3.19) is also always positive; it appears in (3.14) as a negative component of the second term. The curvature of the objective function is mainly dependent on the values of \(U_1\) and \(U_2\). The more likely event is the occurrence of a negative hedging error, for both \(U_1\) and \(U_2\). In the worst case, the hedging error in the second time period is higher than that in the first time period; this suggests that \(U_2\), which is likely to be highly negative, may result in a negative diagonal element. In this situation, some of the eigenvalues of the Hessian would be negative and the solution may have a mid-range element.
3.3 Illustrations

The illustrations show the same output variables as outlined in Chapter 3 including the Hessians of the objective function at the solution and the minimax hedging error for the two time periods, $U_1(n_{1r}, S_{1r}) + U_2(n_{2r}, S_{2r})$. The first three illustrations show the performance of the algorithm when applied to three different options. The last illustration shows the performance of the algorithm when applied to the portfolio comprising the three options. The optimality of the solution for the portfolio case is discussed in Section 5.

In the illustrations, we use data from the empirical study in Chapter 4. The rebalancing date is December 4, 1991. The expiration date for all three options is August 1992. Their exercise prices, as well as other parameters, are given in Section 6 of Chapter 4.
Illustration 3.1

Hedging the risk of writing an option on the Guinness Stock:

\[ n_t \in \mathbb{R}^1, \quad S \in \mathbb{R}^2 \]

Objective function:

\[ f(n_t, S) \]

subject to

\[ 4.731038e+02 \leq S_{t+1} \leq 5.301120e+02 \]
\[ 4.637506e+02 \leq S_{t+2} \leq 5.447065e+02 \]

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Squared Penalty Term</th>
<th>Non-Squared Penalty Term</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-06</td>
<td>1.000000 e-06</td>
<td>1.000000 e-08</td>
<td></td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
<td></td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
<td></td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
<td></td>
</tr>
<tr>
<td>( n_t )</td>
<td>4.807649 e+01</td>
<td>4.807649 e+01</td>
<td>4.807649 e+01</td>
<td></td>
</tr>
<tr>
<td>( (S_{t+1}, S_{t+2})^T )</td>
<td>4.990000 e+02</td>
<td>4.990000 e+02</td>
<td>4.990000 e+02</td>
<td></td>
</tr>
<tr>
<td>( (S_{t+1}, S_{t+2})^T )</td>
<td>4.990000 e+02</td>
<td>4.990000 e+02</td>
<td>4.990000 e+02</td>
<td></td>
</tr>
<tr>
<td>( ldl )</td>
<td>7.766443 e-04</td>
<td>7.901457 e-04</td>
<td>1.004980 e-04</td>
<td></td>
</tr>
<tr>
<td>( \Psi )</td>
<td>-2.053620 e-07</td>
<td>-2.063732 e-07</td>
<td>-9.967024 e-09</td>
<td></td>
</tr>
<tr>
<td>( \Phi(n_{t*}) )</td>
<td>1.800928 e+05</td>
<td>1.800928 e+05</td>
<td>1.800928 e+05</td>
<td></td>
</tr>
<tr>
<td>( (S_{t+1}, S_{t+2})^T )</td>
<td>4.723716 e+01</td>
<td>4.723717 e+01</td>
<td>4.726027 e+01</td>
<td></td>
</tr>
<tr>
<td>( (S_{t+1}, S_{t+2})^T )</td>
<td>4.971921 e+02</td>
<td>4.971881 e+02</td>
<td>4.972337 e+02</td>
<td></td>
</tr>
<tr>
<td>( k_{\alpha_k} )</td>
<td>1 ( \forall k &gt; k_{\alpha} )</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>No. of iterations</td>
<td>3</td>
<td>3</td>
<td>607</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>0.2 sec</td>
<td>0.2 sec</td>
<td>43.4 sec</td>
<td></td>
</tr>
<tr>
<td>Eigenvalues of numerically evaluated ( Hessian_{nn} )</td>
<td>1.213150 e+04</td>
<td>1.213190 e+04</td>
<td>1.212740 e+04</td>
<td></td>
</tr>
<tr>
<td>Eigenvalues of numerically evaluated ( Hessian_{SS} )</td>
<td>-2.345774 e+02</td>
<td>-2.345745 e+02</td>
<td>-2.340270 e+02</td>
<td></td>
</tr>
<tr>
<td>Eigenvalues of Hessian approximation ( Hessian_{BFGS} )</td>
<td>6.091426 e+02</td>
<td>6.091427 e+02</td>
<td>6.078027 e+02</td>
<td></td>
</tr>
<tr>
<td>Minimax Hedging Error</td>
<td>-585 points</td>
<td>-585 points</td>
<td>-585 points</td>
<td></td>
</tr>
</tbody>
</table>
Illustration 3.2

Hedging the risk of writing an option on the Prudential Stock:

\[ n_r \in \mathbb{R}^1, \quad S \in \mathbb{R}^2 \]

Objective function:

\[ f(n_r, S) \]

subject to

\[ 2.142715e + 02 \leq S_{t+1} \leq 2.400908e + 02 \]
\[ 2.100345e + 02 \leq S_{t+2} \leq 2.467007e + 02 \]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Proposed Algorithm</th>
<th>Kiwiels’s Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
<tr>
<td>( n_r )</td>
<td>8.021090 e+01</td>
<td>8.021090 e+01</td>
</tr>
<tr>
<td>((S_{t+1}, S_{t+2})^T)</td>
<td>2.260000 e+02</td>
<td>2.260000 e+02</td>
</tr>
<tr>
<td>( ldl )</td>
<td>1.963846 e-05</td>
<td>1.370546 e-05</td>
</tr>
<tr>
<td>( \Psi )</td>
<td>-2.527874 e-13</td>
<td>-1.254621 e-13</td>
</tr>
<tr>
<td>( \Phi(n_r) )</td>
<td>2.585991 e+04</td>
<td>2.585991 e+04</td>
</tr>
<tr>
<td>( n_r^* )</td>
<td>7.816638 e+01</td>
<td>7.816637 e+01</td>
</tr>
<tr>
<td>((S_{t+1}^<em>, S_{t+2}^</em>)^T)</td>
<td>2.302661 e+02</td>
<td>2.302661 e+02</td>
</tr>
<tr>
<td>( k_{G} )</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>No. of iterations</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Time</td>
<td>0.5 sec</td>
<td>0.5 sec</td>
</tr>
</tbody>
</table>

Eigenvalues of numerically evaluated \( Hessian_{nn} \)

\[ 2.485399 e+03 \]
\[ -1.371005 e+02 \]
\[ 1.309824 e+03 \]

Eigenvalues of numerically evaluated \( Hessian_{SS} \)

\[ 2.490262 e+03 \]
\[ -1.381234 e+02 \]
\[ 6.033050 e+02 \]

Eigenvalues of Hessian approximation \( Hessian_{BFGS} \)

\[ 1.337187 e+03 \]

Minimax Hedging Error

\[ -230 \text{ points} \]
\[ -230 \text{ points} \]
\[ -232 \text{ points} \]
Illustration 3.3

Hedging the risk of writing an option on the British Telecom Stock:

\[ n_t \in \mathbb{R}^1, \quad S \in \mathbb{R}^2 \]

Objective function:

\[ f(n_t, S) \]

subject to

\[ 3.289920e+02 \leq S_{t+1} \leq 3.686350e+02 \]
\[ 3.224879e+02 \leq S_{t+2} \leq 3.787839e+02 \]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Proposed Algorithm</th>
<th>Kiwiell's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>Squared Penalty Term</td>
<td>1.000000 e-06</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
</tbody>
</table>

\[ n_t \]

\[ (S_{t+1}, S_{t+2})^T \]

\[ |dl| \]

\[ \Psi \]

\[ \Phi(n_{t*}) \]

\[ n_{t*} \]

\[ (S_{t+1*}, S_{t+2*})^T \]

\[ k_\alpha | \alpha_k = 1; \forall k > k_\alpha \]

No. of iterations | 12 | 13 | 240

Time | 1.3 sec | 1.3 sec | 17.1 sec

Eigenvalues of numerically evaluated Hessian\( nn \)

\[ 7.576590 \text{ e+03} \]

Eigenvalues of numerically evaluated Hessian\( SS \)

\[ 1.373963 \text{ e+02} \]

Eigenvalues of Hessian approximation Hessian\( BFGS \)

\[ 1.783625 \text{ e+03} \]

Minimax Hedging Error

-318 points
Illustration 3.4

Hedging the risk of writing three options: Guinness, Prudential and British Telecom:

\[ n_t \in \mathbb{R}^3, \quad S \in \mathbb{R}^6 \]

Objective function:

\[ f(n_t, S) \]

subject to

\[
\begin{align*}
4.731038e+02 + 02 & \leq S_{t+1}^1 & \leq 5.301120e + 02 \\
2.142715e + 02 & \leq S_{t+1}^2 & \leq 2.400908e + 02 \\
3.289920e + 02 & \leq S_{t+1}^3 & \leq 3.686350e + 02 \\
4.637506e + 02 & \leq S_{t+2}^1 & \leq 5.447065e + 02 \\
2.100345e + 02 & \leq S_{t+2}^2 & \leq 2.467007e + 02 \\
3.224879e + 02 & \leq S_{t+2}^3 & \leq 3.787839e + 02
\end{align*}
\]

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwiell's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-05</td>
<td>1.000000 e-05</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
</tbody>
</table>

\[ n_t \]

\[ (S_{t+1}, S_{t+2})^T \]

\[ ldl \]

\[ \Psi \]

\[ \Phi(n_{t^*}) \]
<table>
<thead>
<tr>
<th>Proposed Algorithm</th>
<th>Kiwiel’s Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>$n_t^*$</td>
<td>$n_t^*$</td>
</tr>
<tr>
<td>4.867006 e+01</td>
<td>4.870550 e+01</td>
</tr>
<tr>
<td>7.576715 e+01</td>
<td>7.577643 e+01</td>
</tr>
<tr>
<td>6.343740 e+01</td>
<td>6.343814 e+01</td>
</tr>
</tbody>
</table>

$\left(S_{t+1}^*, S_{t+2}^*\right)^T$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.002876 e+02</td>
<td>5.003543 e+02</td>
<td>4.997487 e+02</td>
</tr>
<tr>
<td>2.271383 e+02</td>
<td>2.271511 e+02</td>
<td>2.271341 e+02</td>
</tr>
<tr>
<td>3.495103 e+02</td>
<td>3.495164 e+02</td>
<td>3.495293 e+02</td>
</tr>
<tr>
<td>5.447065 e+02</td>
<td>5.447065 e+02</td>
<td>5.447065 e+02</td>
</tr>
<tr>
<td>2.100354 e+02</td>
<td>2.100354 e+02</td>
<td>2.100354 e+02</td>
</tr>
<tr>
<td>3.224879 e+02</td>
<td>3.224879 e+02</td>
<td>3.224879 e+02</td>
</tr>
</tbody>
</table>

$k_{\alpha} | \alpha_k = 1; \forall k > k_{\alpha}$

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>21</td>
</tr>
<tr>
<td>5.5 sec</td>
<td>13.8 sec</td>
</tr>
</tbody>
</table>

Remark

Eigenvalues of numerically evaluated

$Hessian_{nn}$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.243037 e+03</td>
<td>2.243432 e+03</td>
<td>2.240437 e+03</td>
</tr>
<tr>
<td>5.401405 e+03</td>
<td>5.402402 e+03</td>
<td>5.430308 e+03</td>
</tr>
<tr>
<td>1.214053 e+04</td>
<td>1.213337 e+04</td>
<td>1.220856 e+04</td>
</tr>
</tbody>
</table>

Eigenvalues of numerically evaluated

$Hessian_{SS}$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-6.554224 e+02</td>
<td>-6.541567 e+02</td>
<td>-6.642093 e+02</td>
</tr>
<tr>
<td>-5.124750 e+02</td>
<td>-5.116078 e+02</td>
<td>-5.192260 e+02</td>
</tr>
<tr>
<td>-3.777584 e+02</td>
<td>-3.770944 e+02</td>
<td>-3.830674 e+02</td>
</tr>
<tr>
<td>5.006440 e+02</td>
<td>4.995139 e+02</td>
<td>5.392193 e+02</td>
</tr>
<tr>
<td>9.704352 e+02</td>
<td>9.690472 e+02</td>
<td>1.050029 e+03</td>
</tr>
<tr>
<td>1.829004 e+03</td>
<td>1.827328 e+03</td>
<td>1.747872 e+03</td>
</tr>
</tbody>
</table>

Eigenvalues of Hessian approximation

$Hessian_{BFGS}$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.347986 e+03</td>
<td>8.316125 e+03</td>
<td>1.347986 e+03</td>
</tr>
<tr>
<td>1.469045 e+06</td>
<td>1.343950 e+06</td>
<td>1.469045 e+06</td>
</tr>
<tr>
<td>3.275932 e+07</td>
<td>1.837951 e+09</td>
<td>3.275932 e+07</td>
</tr>
</tbody>
</table>

Minimax Hedging Error

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-1031 points</td>
</tr>
</tbody>
</table>
In this section, we present CAPM minimax as a risk management policy and take a particular rebalancing date to illustrate the application of the algorithm in order to implement this policy. This strategy defines the worst case scenario in terms of the two components of total volatility: the market risk and the specific risk. We used the Capital Asset Pricing Model (CAPM) as a basis for a price determination function, and developed a CAPM-based minimax hedging strategy that uses both components of total volatility.

The minimax problem:

\[
\begin{align*}
\min_{n_t, E_t} & \quad f(n_t, E_{t+1}) \\
\text{subject to} & \quad E_t \leq E_{t+1} \leq E_u
\end{align*}
\]  

4.1 The objective function of CAPM minimax

We define \( U_1 : \mathbb{R}^k \times \mathbb{R}^{k+1} \rightarrow \mathbb{R}^1 \), \( U_2 : \mathbb{R}^k \rightarrow \mathbb{R}^t \), \( U : \mathbb{R}^k \times \mathbb{R}^{k+1} \rightarrow \mathbb{R}^{k+1} \), \( n_t \in \mathbb{R}^k \), \( E_{t+1} \in \mathbb{R}^{k+1} \) and \( Q \) as a \((k + 1) x (k + 1)\) positive definite diagonal weighting matrix. \( U_1 \) refers to the potential hedging error for the next period. \( U_2 \) refers to transaction costs associated with buying or selling of stocks. \( U^d \in \mathbb{R}^{k+1} \) is the vector of desired values for the potential hedging error for the two periods and the transaction cost terms: we use a desired value of zero, i.e. the desired hedging error is zero and the desired transaction cost is zero.

The objective function is given by

\[
f(n_t, E_{t+1}) = \frac{1}{2} < U - U^d, Q(U - U^d) >
\]

where
\[ n_t = \begin{bmatrix} n^1_t \\ \vdots \\ n^k_t \end{bmatrix} \quad \text{and} \quad E_{t+1} = \begin{bmatrix} M^1_{t+1} \\ \vdots \\ M^k_{t+1} \end{bmatrix} \] (4.4)

\[ U(n_t, E_{t+1}) = \begin{bmatrix} U_1(n_t, E_{t+1}) \\ \vdots \\ U_k(n_t) \end{bmatrix} \quad \text{and} \quad U^d = \begin{bmatrix} U^d_1 \\ \vdots \\ U^d_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \] (4.5)

\[ U_1(n_t, E_{t+1}) = \sum_{i=1}^k n^i_t (S^i_{t+1}(E_{t+1}) - S^i_t) + \sum_{i=1}^k N^i (B^i_t - B^i_{t+1}(S^i_{t+1}(E_{t+1}))) 
+ \sum_{i=1}^k (-n^i_t - n^i_{t-1}) S^i_t + C^i_{t-1}(1 + r\Delta t) r\Delta t \] (4.6)

\[ U_2(n_t) = \begin{bmatrix} U^1_2(n^1_t) \\ \vdots \\ U^k_2(n^k_t) \end{bmatrix} \] (4.7)

where \( U^i_2(n^i_t) = \hat{R}(n^i_t - n^i_{t-1}) S^i_t \). (4.8)

### 4.2 The Hessians

In this section, we distinguish between three Hessians: \textit{Hessian}_{nn} which is the Hessian with respect to \( n_t \); \textit{Hessian}_{EE} which is the Hessian with respect to \( E_{t+1} \); and \textit{Hessian}_{BFGS} which is the BFGS approximation to the Hessian of the max-function. The remarks in Section 2.2 apply.

Eqns (2.9) to (2.12) are applicable to the Hessians of the objective function for CAPM minimax with one change of variable: the maximizing variable should be \( E_{t+1} \). However, Eqns(2.13) and (2.15) can not be adapted by a simple change of variable; there is no existing option pricing model that relates the value of an option to a market index and to specific error variables. Because of this, we have no analytical expression for \( \frac{\partial^2 g_{t+1}}{\partial (E_{t+1})^2} \). As a consequence, we can not evaluate \textit{Hessian}_{EE}. 

---

224
4.3 Illustration

The illustration shows the same output variables as outlined in Chapter 3 including the Hessians of the objective function at the solution and the minimax hedging error $U_t(n_t, E_{t+1})$. The proposed algorithm found a mid-range solution; this solution is consistent with the solution found by another minimax algorithm. As noted above, we can not evaluate $Hessian_{EE}$ and present it as evidence of the concavity of the objective function in the $E_{t+1}$-space. However, we have evaluated the objective function at all extreme points. We present the highest value of the objective function $f(n_t, \overline{E}_{t+1})$ achieved at an extreme point, represented by $\overline{E}_{t+1}$; this value is less than $\Phi(n_t)$. This suggests that the objective function is concave in the $E_{t+1}$-space at the solution.

We illustrate the performance of the proposed algorithm under CAPM minimax using a portfolio of 5 options from the simulation study in Chapter 6. The parameters used in the simulation apply to this particular portfolio.
Illustration 4.1: extreme point solution

Hedging the risk of holding a portfolio of five written call options:

\[ n_r \in \mathbb{R}^5, \quad E_{t+1} \in \mathbb{R}^6 \]

Objective function:

\[ f(n_r, E_{t+1}) \]

subject to

\[
egin{align*}
2.972739e+03 & \leq M_{t+1} \leq 3.087715e+03 \\
-2.360554e-02 & \leq \varepsilon^1_{t+1} \leq 2.360554e-02 \\
-4.770382e-02 & \leq \varepsilon^2_{t+1} \leq 4.770382e-02 \\
-3.302775e-02 & \leq \varepsilon^3_{t+1} \leq 3.302775e-02 \\
-3.302775e-02 & \leq \varepsilon^4_{t+1} \leq 3.302775e-02 \\
-4.475385e-02 & \leq \varepsilon^5_{t+1} \leq 4.475385e-02
\end{align*}
\]

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epsilon ( \varepsilon )</td>
<td>1.000000 e-08</td>
<td>1.000000 e-08</td>
</tr>
<tr>
<td>Lambda ( \lambda )</td>
<td>5.000000 e-01</td>
<td>5.000000 e-01</td>
</tr>
<tr>
<td>Line coefficient ( c )</td>
<td>1.000000 e-04</td>
<td>1.000000 e-04</td>
</tr>
<tr>
<td>Penalty coefficient ( C )</td>
<td>1.000000 e+06</td>
<td>1.000000 e+06</td>
</tr>
</tbody>
</table>

\[ n_r \]

<table>
<thead>
<tr>
<th>( n_r )</th>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.893995 e+01</td>
<td>3.893995 e+01</td>
<td>3.893995 e+01</td>
</tr>
<tr>
<td>6.701772 e+01</td>
<td>6.701772 e+01</td>
<td>6.701772 e+01</td>
</tr>
<tr>
<td>3.724842 e+01</td>
<td>3.724842 e+01</td>
<td>3.724842 e+01</td>
</tr>
<tr>
<td>3.002542 e+01</td>
<td>3.002542 e+01</td>
<td>3.002542 e+01</td>
</tr>
<tr>
<td>3.576585 e+01</td>
<td>3.576585 e+01</td>
<td>3.576585 e+01</td>
</tr>
</tbody>
</table>

\[ E_{t+1} \]

<table>
<thead>
<tr>
<th>( E_{t+1} )</th>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.915059 e+03</td>
<td>2.915059 e+03</td>
<td>2.915059 e+03</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

\[ ldl \]

<table>
<thead>
<tr>
<th>( ldl )</th>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.552360 e-03</td>
<td>2.552360 e-03</td>
<td>1.007951 e-03</td>
</tr>
<tr>
<td>( \Phi(n_{t+1}) )</td>
<td>2.211716 e+06</td>
<td>2.211741 e+06</td>
</tr>
<tr>
<td>( f(n_r, E_{t+1}) )</td>
<td>2.096005 e+06</td>
<td>2.096005 e+06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( n_{t+1} )</th>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.433818 e+01</td>
<td>5.433818 e+01</td>
<td>5.433818 e+01</td>
</tr>
<tr>
<td>7.048898 e+01</td>
<td>7.048898 e+01</td>
<td>7.048898 e+01</td>
</tr>
<tr>
<td>4.677316 e+01</td>
<td>4.677316 e+01</td>
<td>4.677316 e+01</td>
</tr>
<tr>
<td>4.887535 e+01</td>
<td>4.887535 e+01</td>
<td>4.887535 e+01</td>
</tr>
<tr>
<td>4.885359 e+01</td>
<td>4.885359 e+01</td>
<td>4.885359 e+01</td>
</tr>
<tr>
<td>Proposed Algorithm</td>
<td>Kiwiels's Algorithm</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
<td></td>
</tr>
<tr>
<td>$E_{1,1}$</td>
<td>$E_{1,1}$</td>
<td></td>
</tr>
<tr>
<td>3.036498 e+03</td>
<td>3.036498 e+03</td>
<td></td>
</tr>
<tr>
<td>2.360554 e-02</td>
<td>2.360554 e-02</td>
<td></td>
</tr>
<tr>
<td>-1.189297 e-03</td>
<td>-1.189297 e-03</td>
<td></td>
</tr>
<tr>
<td>-1.854275 e-02</td>
<td>-1.854275 e-02</td>
<td></td>
</tr>
<tr>
<td>3.302775 e-02</td>
<td>3.302775 e-02</td>
<td></td>
</tr>
<tr>
<td>3.462539 e-02</td>
<td>3.462539 e-02</td>
<td></td>
</tr>
</tbody>
</table>

$k_{\alpha} l_{\alpha} = 1; \forall k > k_{\alpha}$

<table>
<thead>
<tr>
<th>No. of iterations</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>32.6 sec</td>
</tr>
<tr>
<td>53</td>
<td>32.6 sec</td>
</tr>
<tr>
<td>268</td>
<td>122.4 sec</td>
</tr>
</tbody>
</table>

Eigenvalues of numerically evaluated $Hessian_{nn}$

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Eigenvalues</th>
<th>Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.768598 e+04</td>
<td>2.768598 e+04</td>
<td>2.768678 e+04</td>
</tr>
<tr>
<td>2.992023 e+04</td>
<td>2.992023 e+04</td>
<td>2.991883 e+04</td>
</tr>
<tr>
<td>3.203914 e+04</td>
<td>3.203914 e+04</td>
<td>3.204169 e+04</td>
</tr>
<tr>
<td>4.144918 e+04</td>
<td>4.144918 e+04</td>
<td>4.146307 e+04</td>
</tr>
<tr>
<td>5.099510 e+04</td>
<td>5.099510 e+04</td>
<td>5.094965 e+04</td>
</tr>
</tbody>
</table>

Eigenvalues of numerically evaluated $Hessian_{EE}$ (not available)

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Eigenvalues</th>
<th>Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Eigenvalues of Hessian approximation $Hessian_{BFGS}$

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Eigenvalues</th>
<th>Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.002777 e+06</td>
<td>1.002777 e+06</td>
<td>n.a.</td>
</tr>
<tr>
<td>2.294899 e+05</td>
<td>2.294899 e+05</td>
<td>n.a.</td>
</tr>
<tr>
<td>8.421676 e+05</td>
<td>8.421676 e+05</td>
<td>n.a.</td>
</tr>
<tr>
<td>6.307115 e+05</td>
<td>6.307115 e+05</td>
<td>n.a.</td>
</tr>
<tr>
<td>3.070709 e+05</td>
<td>3.070709 e+05</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Minimax Hedging Error

<table>
<thead>
<tr>
<th>Minimax Hedging Error</th>
<th>Minimax Hedging Error</th>
<th>Minimax Hedging Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1629 points</td>
<td>1629 points</td>
<td>1626 points</td>
</tr>
</tbody>
</table>
5 Optimality of the minimax solution

We compare the optimality of the minimax solution for the portfolio case to its optimality for the individual cases. This is to illustrate that a better solution can be achieved by optimizing over the whole portfolio as compared to optimizing over the individual parts. We note that this better solution is from the point of view of optimization; this is not necessarily consistent with better performance from the point of view of risk management.

We compare the performance, for the portfolio case and for the individual cases, of the minimax algorithm within the context of Two-Period minimax. The objective function value is one possible measure of performance. However, the objective function value in the portfolio case is not directly comparable to that of the summation over individual cases. A more suitable measure is the absolute value of the Two-Period minimax hedging error. For the portfolio case, the algorithm's performance is given the absolute value of its minimax hedging error. For the individual cases, it is measured by summing over the minimax hedging errors for the individual options. The lower the absolute value of the performance for the portfolio of options, the better is the performance of the algorithm when applied to portfolios.

From Illustrations 3.1, 3.2 and 3.3, we calculated the sum over the absolute values of the minimax hedging errors for all three options. From Illustration 3.4, we extracted the minimax hedging error for the portfolio. Table 5 gives the performance of the algorithm when it was applied to the portfolio and when it was applied to the three options individually. We also give the values for the alternative algorithm.
Table 5  Performance when algorithms were applied to a portfolio of three options and when they were applied to the individual options.

<table>
<thead>
<tr>
<th></th>
<th>Proposed Algorithm</th>
<th>Kiwiel's Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Squared Penalty Term</td>
<td>Non-Squared Penalty Term</td>
</tr>
<tr>
<td>Minimax hedging error (absolute value)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portfolio of 3 options</td>
<td>10.31 e+02</td>
<td>10.31 e+02</td>
</tr>
<tr>
<td>Summation over 3 individual options</td>
<td>11.34 e+02</td>
<td>11.34 e+02</td>
</tr>
</tbody>
</table>

The proposed algorithm performed better for the portfolio case than for the individual cases. This suggests that the interaction between the variables for the portfolio case has a positive effect on the solution.

6 Conclusion

We have presented a number of illustrations showing the performance of the proposed quasi-Newton algorithm when applied to the hedging problem. The performance of the proposed algorithm, from the point of view of optimization, is consistent with its performance when applied to the test problems in Chapter 3. The convergence of the proposed algorithm is consistent with its theoretical superlinear convergence property.

We have demonstrated that the continuous minimax formulation of the hedging problem may require mid-range solutions. The existence of mid-range solutions imply that solving a discrete minimax formulation of the continuous minimax problem may not result in a true minimax solution.
1 The proposed quasi-Newton algorithm for continuous minimax

1.1 Description of the proposed algorithm and numerical experiments

1.1.1 We developed a quasi-Newton algorithm to solve continuous minimax problems. This algorithm uses a quadratic approximation to the objective function with a second order term that is conditioned by a Hessian approximation. The Hessian we approximated is of the max-function and not of the objective function. The quadratic approximation is augmented with a penalty term to ensure that a direction of descent is found and that the choice of a maximizer corresponds to the minimum-norm subgradient. The use of the penalty term results in only one maximization of the quadratic approximation per iteration; this accounts in part for the simple structure of the algorithm.

1.1.2 We used the proposed algorithm to solve several test problems: convex-concave and convex-convex continuous minimax problems, and discrete minimax problems. We checked the solutions found by the proposed algorithm by using another minimax algorithm or, where appropriate, by using a nonlinear programming algorithm.
1.1.3 We implemented the algorithm and formulated several minimax strategies for risk management in finance. We applied those strategies to the problem of hedging the risk of writing call options. We discuss the application in Section 2.

1.2 Theoretical and numerical results

1.2.1 We have established the global and local convergence properties of the proposed quasi-Newton algorithm and have shown that it converges superlinearly. We have also shown that the maximizer of the quadratic approximation augmented with a penalty term corresponds to the minimum-norm subgradient.

1.2.2 The solutions found by the algorithm were very similar to the solutions found using another minimax algorithm or, where appropriate, using a nonlinear programming algorithm. The proposed algorithm performed comparably well with the other algorithms. In general, the performance of the proposed algorithm is consistent with its theoretical superlinear convergence property.

1.2.3 In the applications to hedging, we have shown that the curvature of the max-function is determined by the second derivative of the option pricing model with respect to the underlying stock. We have also shown that the minimax hedging error may be characterized by a mid-range solution; this implies that alternative algorithms such as those for discrete minimax or for nonlinear programming would fail to find the true minimax solution.
1.3 Future directions

We propose two major extensions to the algorithm:

1. include steps to cope with max-function type constraints, or max-constraints. A Lagrangian function may be augmented with a similar penalty term that will limit the choice of maximizer.

2. incorporate a global optimization algorithm for finding the maximum. This extension would involve adapting the algorithm to problems with special structures. For example, the algorithm may be made more efficient for problems with a maximization subproblem, say, with bounds constraints only, that are known to have corner or extreme point solutions. For large scale optimization, it may be necessary to augment the minimax algorithm with an efficient way of finding the solution to the maximization subproblem. Such a refinement would be important because the number of extreme points to consider is equal to $2^n$ where $n$ is the dimension in the y-space. Also, for large dimensions, a reformulation of the continuous minimax problem into discrete minimax may not provide a solution because the number of constraints, which is equal to the number of extreme points, becomes increasingly difficult to handle for discrete minimax algorithms. With regard to the applications in risk management of very large portfolios, the algorithm should be extended to include steps that will efficiently solve the maximization subproblem.

\footnote{For example, including a global algorithm for concave minimization (see Rosen[36], and Falk and Hoffman[15]).}
2. Applications of the proposed algorithm to risk management in finance

2.1 The development of different hedging strategies

2.1.1 We applied the algorithm to a risk management problem in finance. The applications of the algorithm include the development of a minimax hedging strategy that a hedger dealing with written call options may use in order to manage his risk; our study applied to European options and to American options where there are no dividends. The strategy uses the algorithm as a computation tool to find the minimax hedging error that sets a bound on the absolute value of the actual hedging error.

2.1.2 We developed seven variants of the minimax hedging strategy, three of which were designed to control transaction costs. Two variants included a penalty term for transaction costs in their objective function; the third used a heuristic weighting system to dampen the variation in hedge recommendations, thereby controlling transaction costs. We tested hypotheses about some of the characteristics of the time series of the underlying stock relative to the exercise price under which the strategy based on the algorithm performed best.

2.1.3 Through a simulation we attempted to identify suitable combinations of the weighting coefficient $q$ and transaction cost $K$ that would give the best performance for the variants with transaction costs terms in their objective function.

2.1.4 We explored the feasibility of extending the minimax hedging strategy into a multi-period setting. We considered two multi-period strategies: Two-period minimax and Variable minimax. Two-Period minimax is designed for the more cautious hedger who wishes to have the hedge recommendation for one time period based on the potential hedging error for both that time period and the following one. In contrast, Variable minimax is designed for the more active hedger who wishes to monitor actual hedging
errors on a regular basis for periods shorter than the time period used as the rebalancing interval, and to rebalance before the end of that interval if he considers the progression of actual hedging errors to be unfavourable to him.

2.1.5 Because the performance of the minimax strategy when it is applied to a portfolio of options is inferior to its performance when applied to the options individually, we explored the feasibility of redefining the minimax hedging strategy to make it responsive to market conditions, i.e. redefining the worst case scenario in terms of movements in a market index and in terms of variation in the specific returns of the underlying stock. We considered a new specification of the worst case scenario that would be more plausible, and less extreme, than the compounded scenario of worst cases based on stock price movements. As part of this exploration, we developed a new minimax strategy, CAPM minimax, based on the Market Model, which uses information on the correlation of the underlying stocks with the market index, thereby incorporating the possibility of a broadbased market movement that may push CAPM minimax in an appropriate direction in its search for a solution.

2.2 Results from the simulation and empirical studies

2.2.1 In both the simulation and the limited empirical study of the seven variants of the minimax strategy, two variants performed slightly less well than delta hedging, two variants performed much the same as delta hedging, and the three variants that constrained transaction costs, either in the objective function or heuristically, were found to be robust in the sense that they consistently outperformed delta hedging, the benchmark strategy, for a wide range of moneyness. In the following we refer to these as the robust variants.
2.2.2 The tests on several hypotheses establish that the robust variants performed best on time series of stock prices with several crossovers and abrupt changes. These results suggest that the robust variants are suitable for hedging at-the-money options.

2.2.3 We have found suitable combinations of the weighting coefficient $q$ and transaction cost $K$ that result in the highest performance for variants with transaction cost terms in their objective function. We have also found that, even for a low $K$ value, there is a corresponding $q$ that would still give a better performance for these variants compared to delta hedging.

2.2.4 We have demonstrated the feasibility of a multi-period minimax strategy. In our analysis, Variable minimax outperforms Basic minimax, and Basic minimax outperforms Two-Period minimax.

2.2.5 We have demonstrated the feasibility of a minimax strategy applied to a portfolio of options. CAPM minimax applied to a portfolio of options outperforms the highest performing variant of Basic minimax applied to each individual option in the portfolio, which in turn outperforms CAPM minimax when it is applied to each individual option in the portfolio. This result suggests a distinctive feature of CAPM minimax, its ability to incorporate information about correlations between returns on the underlying stock and returns on the market index, has a positive effect on performance.

2.3 Future directions

The next stage in the development of minimax applied to risk management may include:

1. adapting the three robust variants to make them suitable for hedging written American put options. A minimax strategy may be appropriate for this problem because early exercise of the put option is possible and there are no definite signals that would warn the hedger of an early exercise. In contrast to American put, for American call
options, there may be a definite signal, say the announcement of dividend payments. The possibility of the early exercise of a put option without warning to the hedger may be reformulated as a worst case scenario, and an adaptation of the three robust variants, with this scenario, may make them suitable for this problem. The variants could also be adapted by using existing approximation methods in estimating the value of put options.

2. studying the interaction of transaction cost $K$, weighting coefficient $q$ and rebalancing interval; this would provide some insight on how to decide on a rebalancing interval, given that the actual transaction cost $K$ is known. This study on interaction is important because the performance of variants with transaction cost terms in their objective function may vary as a function of the rebalancing interval for different combinations of $q$ and $K$. For very small rebalancing intervals, trading would be frequent and the performance of the variants may improve; for very large rebalancing intervals, trading would be infrequent and the performance of the variants may worsen.

3. studying the performance of Two-Period minimax under different levels of volatility of the volatility and different lengths of rebalancing intervals. Two-Period minimax can put a bound on the actual hedging error for the second time period; such a bound may become important to a hedger who has failed to rebalance at the end of the first period. In this sense, Two-Period minimax may be an attractive strategy when the volatility of returns on the underlying stock is itself volatile.

4. extending Two-Period minimax to deal with situations where the hedger who has written an American option fails to rebalance at the end of the first period and the option is exercised during the second period. The early exercise of the American option is similar to the occurrence of a worst case scenario for the hedger. The

---

2MacMillan[23] developed an analytic approximation for the American put option.
possibility of early exercise without warning to the hedger may be reformulated as a worst case scenario and a cautious hedger may find such a modified Two-Period minimax strategy potentially suitable within this scenario.

5. studying the performance of Variable minimax for different levels of both volatility and the percentage level; this would give insight into the best combination of volatility and percentage levels to adopt to keep within any levels of acceptability of actual hedging errors.

6. studying the suitability of the market-based CAPM minimax for other hedging problems where the securities are highly sensitive to market movements, e.g. for portfolio insurance problems where stock index futures are used for hedging portfolios of stocks, or the management of bond portfolios.

7. studying the performance of hedgers who use the hedge recommendations from the minimax variants in their rebalancing decisions against the performance of a control group of hedgers who do not use the minimax hedge recommendations. In our simulation and empirical studies we use delta hedging as the benchmark against which we set the performance of the minimax hedging strategies. Although delta hedging is an obvious benchmark to choose, being based on the most widely used model of option pricing, we have not established that hedgers use delta hedging as their main or only strategy. The proposed study would be relevant to a critical assessment of one of our main results, that the robust minimax variants systematically outperform delta hedging, as a contribution to the risk management of derivatives.

8. carrying out a wide ranging empirical study of the performance of minimax strategies as a tool in the risk management of derivative securities for a large number of options in the major international option markets, and over a period of years. The empirical analysis carried out in this study is very limited: a few stock options written in one
exchange for a limited period. Accordingly, we cannot generalize our results from the empirical study: different option markets and the markets for their underlying stocks may have different dynamics, different patterns of volatility, and different degrees of market manipulation; further, hedgers may have culturally bound risk preferences. A wide ranging empirical study would help establish the generalizability of our results.

3 Finale: the relation of continuous minimax to risk management

We have demonstrated that continuous minimax could be a basis for practical hedging strategies in risk management of derivative securities, and we have established that continuous minimax algorithms reach mid-range solutions that neither discrete minimax nor nonlinear programming algorithms can reach. In this sense, we have demonstrated that continuous minimax could uniquely contribute to the development of efficient risk management.
REFERENCES


