Mathematical Programs with Equilibrium Constraints

ZHI-QUAN LUO,
JONG-SHI PANG,
AND DANIEL RALPH

This book provides a solid foundation and an extensive study for an important class of constrained optimization problems known as Mathematical Programs with Equilibrium Constraints (MPEC), which are extensions of bilevel optimization problems. The book begins with the description of many source problems arising from engineering and economics that are amenable to treatment by the MPEC methodology. Error bounds and parametric analysis are the main tools to establish a theory of exact penalization, a set of MPEC constraint qualifications and the first- and second-order optimality conditions. The book also describes several iterative algorithms such as a penalty-based interior point algorithm, an implicit programming algorithm and a piecewise sequential quadratic programming algorithm for MPECs. Results in the book will have significant impacts in such disciplines as engineering design, economics and game equilibria, and transportation planning, within all of which MPEC has a central role to play in the modeling of many practical problems.

A useful resource for applied mathematicians in general, this book will be a particularly valuable tool for operations researchers, transportation, industrial, and mechanical engineers, and mathematical programmers.

Mathematical Programs
with
Equilibrium Constraints

Mathematical Programs with Equilibrium Constraints

ZHI-QUAN LUO

McMaster University

JONG-SHI PANG

The Johns Hopkins University

DANIEL RALPH

The University of Melbourne

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, United Kingdom
40 West 20th Street, New York, NY 10011–4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First Published 1996

Typeset in Computer Modern

Library of Congress Cataloging-in-Publication Data

Luo, Zhi-Quan.

Mathematical programs with equilibrium constraints / Zhi-Quan Luo,
Jong-Shi Pang, Daniel Ralph.

p. cm.

Includes bibliographical references and index.

ISBN 0–521–57290–8

1. Mathematical optimization. 2. Nonlinear programming.
I. Pang, Jong-Shi. II. Ralph, Daniel. III. Title.

QA402.5.L83 1997
519.7′6–dc20 99–19428
CIP

A catalog record for this book is available from the British Library.

ISBN 0–521–57290–8 hardback

Transferred to digital printing 2004

To our families

Contents

Numbering System

The six chapters of the book are numbered from 1 to 6, the sections are denoted by decimal numbers of the type **2.3** (meaning Section 3 of Chapter 2). Many sections are further divided into subsections, some subsections are numbered, others are not. The numbered subsections are by decimal numbers following the section numbers; e.g., Subsection **1.3.1** means Chapter 1, Section 3, Subsection 1.

All definitions, results, and miscellaneous items are numbered consecutively within each section in the form **1.3.5**, **1.3.6**, meaning Items 5 and 6 in Section 3 of Chapter 1. All items are also identified by their types (e.g., **1.4.1 Proposition.**, **1.4.2 Remark.**). When an item is referred to in the text, it is called out as Algorithm **5.2.1**, Theorem **4.1.7**, etc.

Equations are numbered consecutively in each section by (1), (2), etc. Any reference to an equation in the same section is by this number only, whereas equations in another section are identified by chapter, section, and equation. Thus (3.1.4) means Equation (4) in Section 1 of Chapter 3.

Acronyms

AVI	Affine Variational Inequality
BIF	B(ouligand)-Differentiable Implicit Function Condition
CQ	Constraint Qualification
C^r	Continuously differentiable of order r
CRCQ	Constant Rank Constraint Qualification
GBIF	Global BIF Condition
IMP	Implicit Programming
KKT	Karush-Kuhn-Tucker
LCP	Linear Complementarity Problem
LICQ	Linear Independence Constraint Qualification
MFCQ	Mangasarian-Fromovitz Constraint Qualification
MP	Mathematical Program
MPAEC	Mathematical Program with Affine Equilibrium Constraints
MPEC	Mathematical Program with Equilibrium Constraints
NCP	Nonlinear Complementarity Problem
NLP	Nonlinear Program
PCP	Piecewise Programming
PC^r	Piecewise smooth of order r
PIPA	Penalty Interior Point Algorithm
PSQP	Piecewise Sequential Quadratic Programming
SBCQ	Sequentially Bounded Constraint Qualification
SCOC	Strong Coherent Orientation Condition
SMFCQ	Strict Mangasarian-Fromovitz Constraint Qualification
SQP	Sequential Quadratic Programming
SRC	Strong Regularity Condition
VI	Variational Inequality

Glossary of Notation

Scalars

sgn t	the sign, $1, -1, 0$, of a positive, negative, or zero scalar t
$t_+ \equiv \max(0, t)$	the nonnegative part of a scalar
$t_- \equiv \max(0, -t)$	the nonpositive part of a scalar

Spaces

$\Re^n$	real n-dimensional space
$\Re$	the real line
$\Re^{n \times m}$	the space of $n \times m$ real matrices
$\Re^n_+$	the nonnegative orthant of $\Re^n$
$\Re^n_{++}$	the positive orthant of $\Re^n$

Vectors

z^T	the transpose of a vector z
$\{z^k\}$	a sequence of vectors $z^1, z^2, z^3, \ldots$
$x^T y$	the standard inner product of vectors in $\Re^n$
$\|x\| \equiv \sqrt{x^T x}$	the Euclidean norm of a vector $x \in \Re^n$
$x \geq y$	the (usual) partial ordering: $x_i \geq y_i, i = 1, \ldots n$
$x > y$	the strict ordering: $x_i > y_i, i = 1, \ldots, n$
$\min(x, y)$	the vector whose i-th component is $\min(x_i, y_i)$
$\max(x, y)$	the vector whose i-th component is $\max(x_i, y_i)$
$x \circ y \equiv (x_i y_i)$	the Hadamard product of x and y
$x \perp y$	x and y are perpendicular
$z^+ \equiv \max(0, z)$	the nonnegative part of a vector z
$z^- \equiv \max(0, -z)$	the nonpositive part of a vector z

Matrices

$\det A$	the determinant of a matrix A
A^{-1}	the inverse of a matrix A
$\|A\|$	the Euclidean norm of a matrix A
A^T	the transpose of a matrix A
A_α	the columns of A indexed by α
$A_{\alpha\cdot}$	the rows of A indexed by α
I	the identity matrix of appropriate order
I_k	the identity matrix of order k
$\mathrm{diag}(a)$	the diagonal matrix with diagonal elements equal to the components of the vector a

Functions

$f : \mathcal{D} \to \mathcal{R}$	a mapping with domain $\mathcal{D}$ and range $\mathcal{R}$
$f \circ g$	composition of two functions f and g
∇f	$(\partial f_i/\partial x_j)$, the $m \times n$ Jacobian of a mapping $f : \Re^n \to \Re^m$ $(m \geq 2)$
$\nabla_\beta f_\alpha$	$(\partial f_i/\partial x_j)_{i\in\alpha}^{j\in\beta}$, a submatrix of ∇f
$\nabla\theta$	$(\partial\theta/\partial x_j)$, the gradient of a function $\theta : \Re^n \to \Re$
$\nabla_y g(x, y)$	the partial Jacobian matrix of g with respect to y
$\nabla^2\theta$	Hessian matrix of the scalar-valued function θ
$f'(\cdot\,;\cdot)$	directional derivative of the mapping f
f^{-1}	the inverse of f
$o(t)$	any function such that $\lim_{t\to 0} \frac{o(t)}{t} = 0$
$\Pi_K(x)$	the Euclidean projection of x on the set K
$\inf f(x)$	the infimum of the function f
$\sup f(x)$	the supremum of the function f
$\mathrm{dist}(x, W)$	distance function from vector x to set W
F_C	normal map associated with function F and set C

Sets

$\in$	element membership
$\notin$	not an element of
$\emptyset$	the empty set
$\subseteq$	set inclusion
$\subset$	proper set inclusion
$\cup, \cap, \times$	union, intersection, Cartesian product
$\prod S_i$	Cartesian product of sets S_i
$S_1 \setminus S_2$	the difference of sets S_1 and S_2
$\lvert S \rvert$	the cardinality of a finite set S
∂S	the (topological) boundary of a set S
$\mathrm{cl}\, S$	the (topological) closure of a set S
S^*	the dual cone of S
$0^+ S$	the cone of recession directions of S
$\mathrm{Gr}(\mathcal{A})$	the graph of a multifunction $\mathcal{A}$
$\mathrm{dom}(\mathcal{A})$	the domain of a multifunction $\mathcal{A}$
$\mathbb{B}(x, \delta)$	the closed ball with center at x with radius δ
$\mathrm{argmin}_x f(x)$	the set of x attaining the minimum of the real-valued function $f(x)$
$\mathrm{argmax}_x f(x)$	the set of x attaining the maximum of the real-valued function $f(x)$
$\mathrm{supp}(x)$	the support of vector x
$\mathcal{T}(x; S)$	tangent cone of set S at point $x \in S$
$\mathcal{C}(x; S)$	critical cone of set S at point $x \in S$ relative to an objective function
$\mathcal{N}(x; S)$	normal cone of set S at point $x \in S$
$[a, b]$	a closed interval in $\Re$
(a, b)	an open interval in $\Re$
$x^\perp$	the orthogonal complement of vector x

Problems

AVI (q, M, K)	AVI defined by vector q, matrix M and set K
LCP (q, M)	LCP defined by vector q and matrix M
$\mathrm{SOL}(F, K)$	solution set of the VI (F, K)
$\mathrm{SOL}(q, M, K)$	solution set of the AVI (q, M, K)
VI (F, K)	VI defined by mapping F and set K

MPEC symbols

$v \equiv (\zeta, \pi, \eta)$	MPEC multipliers
$w \equiv (x, y, \lambda)$	variable of MPEC in KKT form, $\lambda \in M(x, y)$
$y(x)$	implicit solution function of lower-level VI
$z \equiv (x, y)$	original MPEC variable
$\mathcal{B}(\bar{x})$	SCOC family of active index sets that define the C^1 pieces of $y(x)$ at $\bar{x}$
$\mathcal{C}(z; \mathcal{F})$	critical cone of MPEC at $z \in \mathcal{F}$ relative to the objective function f $\equiv \bigcup_{\lambda \in M(z)} \mathcal{C}(z, \lambda)$ under full MPEC CQ
$\mathcal{C}(z, \lambda)$	a piece of $\mathcal{C}(z; \mathcal{F})$ corresponding to $\lambda \in M(z)$ $\equiv \mathcal{T}(z; Z) \cap \mathrm{Gr}(\mathcal{LS}_{(z,\lambda)}) \cap \nabla f(z)^{\perp}$
$\mathcal{F}$	MPEC's feasible region given by $Z \cap \mathrm{Gr}(\mathcal{S})$
$\mathcal{F}^{\mathrm{KKT}}$	feasible region of MPEC in KKT form
$\mathcal{I}(x, y)$	set of active indices at $(x, y) \in \mathrm{Gr}(\mathcal{S})$ $\equiv \{i \,:\, g_i(x, y) = 0\}$
$\mathcal{I}_0(x, y, \lambda)$	degenerate index set at $(x, y, \lambda) \in \mathcal{F}^{\mathrm{KKT}}$ $\equiv \{i \,:\, \lambda_i = g_i(x, y) = 0\}$
$\mathcal{I}_+(x, y, \lambda)$	nondegenerate index set at $(x, y, \lambda) \in \mathcal{F}^{\mathrm{KKT}}$ $\equiv \{i \,:\, \lambda_i > g_i(x, y) = 0\}$
$\mathcal{K}(z, \lambda)$	lifted critical cone at $(z, \lambda) \in \mathcal{F}^{\mathrm{KKT}}$
$\mathcal{K}(z, \lambda; dx)$	directional critical set at $(z, \lambda) \in \mathcal{F}^{\mathrm{KKT}}$ along direction dx
$L(x, y, \lambda)$	Lagrangean function for lower-level VI $\equiv F(x, y) + \sum_{i=1}^{\ell} \lambda_i \nabla_y g_i(x, y)$
$\mathcal{L}(z; \mathcal{F})$	MPEC linearized cone at $z \in \mathcal{F}$ $\equiv \mathcal{T}(z, Z) \cap \left(\bigcup_{\lambda \in M(z)} \mathrm{Gr}(\mathcal{LS}_{(z,\lambda)}) \right)$
$\mathcal{L}^{\mathrm{MPEC}}(w, \zeta, \pi, \eta)$	MPEC Lagrangean function
$\mathcal{LS}_{(z,\lambda)}$	linearized solution map at $(z, \lambda) \in \mathcal{F}^{\mathrm{KKT}}$ for lower-level VI; $\mathcal{LS}_{(z,\lambda)}(dx)$ is defined as $\mathrm{SOL}(\nabla_x L(z, \lambda) dx, \nabla_y L(z, \lambda), \mathcal{K}(z, \lambda, dx))$
$\mathcal{LS}^{\mathrm{KKT}}_{(z,\lambda)}(dx)$	set of KKT pairs $(dy, d\lambda)$ of the AVI $(\nabla_x L(z, \lambda) dx, \nabla_y L(z, \lambda), \mathcal{K}(z, \lambda; dx))$

MPEC symbols
(continued)

$M(x, y)$	set of KKT multipliers of VI $(F(x,\cdot), C(x))$ at solution y
$M^c(z; dx)$	set of critical multipliers at $z \in \mathcal{F}$ along direction dx $\equiv \{\lambda \in M(z) : \mathcal{K}(z, \lambda; dx) \neq \emptyset\}$
$M^e(x, y)$	set of extreme points of $M(x, y)$
$\mathcal{S}(x)$	set of rational reactions of lower-level VI $\equiv \text{SOL}(F(x,\cdot), C(x))$
Z	upper-level feasible region of (x, y)

Preface

This monograph deals with a class of constrained optimization problems which we call *Mathematical Programs with Equilibrium Constraints*, or simply, MPECs. Briefly, an MPEC is an optimization problem in which the essential constraints are defined by a parametric variational inequality or complementarity system. The terminology, MPEC, is believed to have been coined in [108]; the word "equilibrium" is adopted because the variational inequality constraints of the MPEC typically model certain equilibrium phenomena that arise from engineering and economic applications. The class of MPECs is an extension of the class of *bilevel programs*, also known as mathematical programs with optimization constraints, which was introduced in the operations research literature in the early 1970s by Bracken and McGill in a series of papers [34, 36, 37]. The MPEC is closely related to the economic problem of Stackelberg game [265] the origin of which predates the work of Bracken and McGill.

Our motivation for writing this monograph on MPEC stems from the practical significance of this class of mathematical programs and the lack of a solid basis for the treatment of these problems. Although there is a substantial amount of previous research on special cases of MPEC, no existing work provides such generality, depth, and rigor as the present study. Our intention in this monograph is to establish a sound foundation for MPEC that we hope will inspire further applications and research on this important problem.

This monograph consists of six chapters. Chapter **1** defines the MPEC, gives a brief description of several source problems, and presents various equivalent formulations of the equilibrium constraints in MPEC; the chapter concludes with some results of existence of optimal solutions. Chapter **2** presents an extensive theory of exact penalty functions for MPEC, using the theory of error bounds for inequality systems. This chapter ends with a brief discussion of how some exact penalty functions formulations of MPEC can be employed to obtain first-order optimality conditions; the latter topic and its extensions are treated in full in the next three chapters. Specifically, Chapter **3** presents the fundamental first-order optimality (i.e., stationarity) conditions of MPEC; Chapter **4** verifies in detail the hypotheses needed for the first-order conditions; Chapter **5** contains results on second-order optimality conditions. The sixth and last chapter presents several algorithms for solving MPECs including an interior point algorithm for MPECs with "monotone" inner problems, a conceptual iterative descent algorithm based on an implicit programming approach, and a locally superlinearly convergent Newton type (sequential quadratic programming) method based on a piecewise programming approach. Some preliminary computational results are reported. The monograph ends with an extensive list of references.

Due to the intrinsic complexity of the MPEC, a comprehensive study of this problem would inevitably require extensive tools from diverse disciplines. Besides a general knowledge of smooth (nonlinear) programming and multivariate analysis, which we assume as prerequisites for this work, such subjects as error bound theory for inequality systems, sensitivity and stability theory for parametric variational inequalities, piecewise smooth analysis, nonsmooth equations, the family of interior point methods, and some basic iterative descent methods for nonlinear programs are all important tools that will be used in this monograph. Since it is not possible for us to review in detail all the background material and keep the monograph within a reasonable length, we have chosen not to organize the preliminary results separately. Instead, we have included only the most useful background results relevant to the topics of discussion.

Throughout the monograph, we have taken several different points of view toward the MPEC, each of which is interesting by itself. Many results obtained herein are new and have not appeared in the literature before. For related approaches and results, we refer to [201, 214, 291, 292, 295]; see also the references in [5, 278].

The general MPEC is a highly nonconvex, nondifferentiable optimization problem that encompasses certain combinatorial features in its constraints. As such, it is computationally very difficult to solve, especially if one wishes to compute a globally optimal solution. Partly due to this pessimistic view, we have not attempted in this monograph to deal with the issue of finding a globally optimal solution to the general problem itself or to its special cases. The algorithms discussed in Chapter **6** are iterative schemes for computing a stationary point of the MPEC (and under mild conditions, a strict local minimum). We refer to [278] for references that discuss some global optimization approaches to solving bilevel programs.

Due to the broad applications of MPEC, this monograph is of interest to readers from diverse disciplines. In particular, operations researchers, economists, design and systems engineers, and applied mathematicians will likely find the subject matter interesting and challenging. We have written the monograph with these individuals in mind.

Acknowledgements

This monograph grew out of a research article [172] authored by us and Professor Shi-Quan Wu of the Academia Sinica in Beijing. We are indebted to Professor Wu's general interest in this subject which has provided the initial stimulus for us to undertake the project. We are also grateful to several funding agencies which have supported our research in the general optimization area. In particular, Tom Luo thankfully acknowledges the Natural Sciences and Engineering Research Council of Canada. Jong-Shi Pang gratefully acknowledges the generous support of the National Science Foundation (U.S.A.) over many years for his research in the general area of variational inequalities and complementarity problems; he is also grateful to the Office of Naval Research for its support on the work in the interior point methods. Danny Ralph wishes to thank the Australian Research Council for its support on this project.

During the preparation of the manuscript, we have presented its contents in several international conferences and received many valuable comments from colleagues. In particular, we have benefited from discussions with Professors Masao Fukushima, Olvi Mangasarian, Jiri Outrata, and Steve Robinson, and Drs. Yang Chen, Eric Ralph, and Houyuan Jiang. Dr. Alan Harvey from the Cambridge University Press has been very encouraging and supportive of this project. We thank them all.

The text of this monograph was typeset by the authors using Leslie Lamport's LaTeX, a document preparation system based on Donald Knuth's

TeX program. We have used the document style files of the book [54] that were prepared by Richard Cottle and Richard Stone, based on the LaTeX book style.

Zhi-Quan (Tom) Luo,	McMaster, Ontario, Canada
Jong-Shi Pang,	Baltimore, Maryland, U.S.A.
Daniel Ralph,	Melbourne, Victoria, Australia

August 19, 1996

1

Introduction

This chapter introduces the main topic of this monograph, that is, the mathematical program with equilibrium constraints. Source problems from engineering and economics are described to justify the need for a thorough study of this important class of optimization problems. The rest of the chapter gives several useful formulations of the equilibrium constraints and presents sufficient conditions for the existence of optimal solutions to these problems.

1.1 Problem Formulation

A Mathematical Program with Equilibrium Constraints (MPEC) is an optimization problem with two sets of variables, $x \in \Re^n$ and $y \in \Re^m$, in which some or all of its constraints are defined by a parametric variational inequality or complementarity system with y as its primary variables and x the parameter vector. More specifically, this problem is defined as follows. Suppose that $f : \Re^{n+m} \to \Re$ and $F : \Re^{n+m} \to \Re^m$ are given functions, $Z \subseteq \Re^{n+m}$ is a nonempty closed set, and $C : \Re^n \to \Re^m$ is a set-valued map with (possibly empty) closed convex values; i.e., for each $x \in \Re^n$, $C(x)$ is a (possibly empty) closed convex subset of $\Re^m$. The set of all vectors $x \in \Re^n$

for which $C(x) \neq \emptyset$ is the *domain* of C and denoted $\mathrm{dom}(C)$. Let X be the projection of Z onto $\Re^n$; i.e.,

$$X = \{x \in \Re^n : (x, y) \in Z \text{ for some } y \in \Re^m\}.$$

The function f is the overall objective function to be minimized; F is the *equilibrium* function of the *inner* problem, Z is a joint upper-level feasible region of the pair (x, y), and $C(x)$ defines the restriction of the variable y for each given $x \in X$. We shall make the blanket assumption that $X \subseteq \mathrm{dom}(C)$. With this setup, the MPEC is:

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & (x,y) \in Z, \text{ and} \\ & y \in \mathcal{S}(x), \end{array} \tag{1}$$

where for each $x \in X$, $\mathcal{S}(x)$ is the solution set of the variational inequality (VI) defined by the pair $(F(x,\cdot), C(x))$; i.e., $y \in \mathcal{S}(x)$ if and only if y is in $C(x)$ and satisfies the inequality:

$$(v - y)^T F(x, y) \geq 0, \quad \text{for all } v \in C(x).$$

In general, the graph of a set-valued map $\mathcal{A} : \Re^n \to \Re^m$ will be denoted $\mathrm{Gr}(\mathcal{A})$; thus

$$\mathrm{Gr}(\mathcal{A}) = \{(x, y) \in \Re^{n+m} : y \in \mathcal{A}(x)\}.$$

By considering the solution map $\mathcal{S}$ as a set-valued map from $\Re^n$ into $\Re^m$, we may write the problem (1) in the compact form:

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & (x,y) \in Z \cap \mathrm{Gr}(\mathcal{S}). \end{array}$$

Let

$$\mathcal{F} \equiv Z \cap \mathrm{Gr}(\mathcal{S}) \tag{2}$$

denote the feasible region of (1). Throughout this monograph, we shall make the blanket assumption that this region is nonempty. We refer the reader to [220] for a comprehensive review of the VI and related problems and to [7] for the fundamental theory of set-valued maps.

The term "equilibrium constraints" in MPEC refers to the variational inequality constraint $y \in \mathcal{S}(x)$. Our usage of the word "equilibrium" reflects the fact that we are particularly interested in the case of MPEC where its constraints represent certain equilibrium applications in engineering and economics that are modeled by variational inequalities.

The formulation (1) of MPEC is very broad and encompasses a large number of interesting special cases. Foremost among these is the case where the mapping $F(x, \cdot)$ is the partial gradient map (with respect to the second argument) of a real-valued C^1 function $\theta : \Re^{n+m} \to \Re$; i.e., $F(x,y) = \nabla_y \theta(x,y)$ for all $(x,y) \in \mathrm{Gr}(C)$ where ∇_y denotes the partial F(réchet)-differentiation with respect to the variable y. In this case, the VI $(F(x,\cdot), C(x))$ is, for each fixed $x \in X$, the set of stationarity conditions of the following optimization problem in the variable y:

$$\begin{array}{ll} \text{minimize} & \theta(x,y) \\ \text{subject to} & y \in C(x). \end{array} \tag{3}$$

This special case of the MPEC has traditionally been known as the *bilevel program* with (3) called its *inner program* or *lower-level program* for an obvious reason. In general, we shall use "argmin" to denote the optimal solution set of a minimization problem. Thus for a given vector x, $\operatorname{argmin}\{\theta(x,y) : y \in C(x)\}$ denotes the optimal solution set of (3). For a convex set $C(x)$, we have

$$\operatorname{argmin}\{\theta(x,y) \,:\, y \in C(x)\} \subseteq \mathcal{S}(x),$$

where $\mathcal{S}(x)$ is the solution set of the VI $(\nabla_y \theta(x,y), C(x))$; moreover, equality holds if in addition $\theta(x,\cdot)$ is convex in the second argument.

The MPEC (1) is a generalization of a bilevel program in which the inner problems are VIs. A bilevel program is in turn a special case of a hierarchical mathematical program which consists of multiple (possibly more than two) levels of optimization. Such multi-level mathematical programs have proven very useful in the modeling of hierarchical decision making processes and in the optimization of engineering designs.

A simple example of a two-level decision making process is as follows. Consider an economic planning process which involves several interacting agents (or individuals). Some agents, collectively called a *leader* or a *principal*, act as superiors who issue directives to the remaining agents, collectively called a *follower* or simply an agent, who act as subordinates to the leader. The leader's directives are described by the variable x and the follower's decision variables are contained in the vector y. The variational inequality constraint $y \in \mathcal{S}(x)$ stipulates that for each of the leader's directives x, the follower will choose a response vector y which is a solution of a decision making problem modeled by the VI $(F(x,\cdot), C(x))$ that depends on x. Based on such *rational* responses from the subordinates, the overall economic planning problem is to determine an optimal vector of the leader's directives x^{opt} along with an equilibrium vector of the follower's responses $y^{\mathrm{equ}} \in \mathcal{S}(x^{\mathrm{opt}})$ in order to minimize an economic performance function modeled by $f(x,y)$ subject to the additional joint feasibility condition $(x,y) \in Z$. With this economic interpretation, the solution set $\mathcal{S}(x)$ for the inner problem is sometimes called the set of *rational reactions* corresponding to x, the solution map $\mathcal{S}$ is called the *reaction map*, and the graph $\mathrm{Gr}(\mathcal{S})$ is called the *rational set*. The objective function $f(x,y)$ and the joint feasible region Z can be used to model additional anticipation of the principal toward the (subordinate) agent's behavior. We will discuss this modeling issue further in the context of the Stackelberg game; see Section **1.2**.

Separately, in the modeling of many engineering design problems as an MPEC, the first-level vector x typically contains the design variables of an engineering process and the second-level vector y contains the state variables of the system; each inner VI $(F(x,\cdot), C(x))$ corresponds to either an optimization or equilibrium problem for a given admissible design x. The overall optimization problem (1) is to determine an optimal pair of design and state variables that will minimize the cost function $f(x,y)$ subject to the joint feasibility condition $(x,y) \in Z$ and the design constraint $y \in \mathcal{S}(x)$. Several design problems of this type will be presented in Section **1.2**.

An important special case of the MPEC (1) is where $C(x)$ is a convex cone in $\Re^m$ for all $x \in X$. In this case, it is known from the theory of variational inequalities [109] that the VI $(F(x,\cdot), C(x))$ is equivalent to a generalized complementarity problem over the cone $C(x)$:

$$y \in C(x), \quad F(x,y) \in C(x)^*, \quad y^T F(x,y) = 0, \tag{4}$$

where for an arbitrary subset S in an Euclidean space $\Re^N$,

$$S^* \equiv \{z \in \Re^N \,:\, z^T v \geq 0 \text{ for all } v \in S\} \tag{5}$$

is the *dual cone* of S. The case $C(x) = \Re^{m_1} \times \Re_+^{m_2}$ for some nonnegative integers m_1 and m_2 such that $m_1 + m_2 = m$ is particularly interesting. In this case, the vectors y and $F(x, y)$ can be partitioned into

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad F(x,y) = \begin{pmatrix} F_1(x,y) \\ F_2(x,y) \end{pmatrix},$$

where y_1, $F_1(x, y) \in \Re^{m_1}$ and y_2, $F_2(x, y) \in \Re^{m_2}$, and the problem (4) becomes

$$\begin{gathered} F_1(x,y) = 0, \\ (y_2, F_2(x,y)) \geq 0, \quad (y_2)^T F_2(x,y) = 0, \end{gathered}$$

which is a mixed, nonlinear complementarity problem. When $m_1 = 0$, the latter problem is a standard nonlinear complementarity problem (NCP).

The LCP (linear complementarity problem) constrained MP (mathematical program) is a special case of the NCP constrained MP in which the function F is linear. The following mathematical program:

$$\begin{array}{ll} \text{minimize} & d^T x + e^T u + f^T v \\ \text{subject to} & Ax + Bu + Cv \geq g, \\ & (u, v) \geq 0, \quad u^T v = 0, \end{array} \tag{6}$$

whose constraints are in the form of a (nonstandard) linear complementarity problem, parametrized by x, has been called a *(linear) complementary program* in [118] and an LPEC in [187]. In turn, this complementary program is a special instance of a *disjunctive program* [14] which is an optimization problem with disjunctive (i.e., "or") constraints. To see that (6) is a disjunctive program, we note that the nonnegativity and complementarity constraint of this problem, $(u, v) \geq 0$, $u^T v = 0$, is equivalent to

$$(u, v) \geq 0, \quad \forall i \ (u_i = 0 \text{ or } v_i = 0),$$

which involves the disjunction "or". In essence, the NCP constrained MP, and more generally, the general MPEC (1) where the inner VIs are formulated in terms of their Karush-Kuhn-Tucker (KKT) conditions (see Subsection **1.3.2**), are special instances of a nonlinear, disjunctive program.

A special case of the linear complementary program is a linear (mixed) integer program in $\{0, 1\}$ variables. Due to this connection, the early study of the former program is closely tied to integer programming; in particular, cutting plane methods [121] and facial techniques [14] have been proposed for solving this problem.

The complexity of MPEC

The computational complexity of a (linear) disjunctive program is well known in the integer programming literature [14]. In essence, this complexity is caused by the disjunctive constraints which lead to some challenging combinatorial issues that typically are the main concern in the design of efficient solution algorithms. As these disjunctive constraints are also present implicitly in a general MPEC, the latter problem can be expected to be quite difficult.

Indeed, the general MPEC (1) is an extremely difficult optimization problem. Besides the intrinsic combinatorial curse of the constraints, the difficulty arises from several other sources that are more akin to this problem considered as a nonlinear program. One such source is the potential lack of convexity and/or closedness of the feasible region $\mathcal{F}$. The special case of the MPEC in which the inner problems are linear programs can be used to elucidate the lack of these useful properties. The following simple numerical example illustrates the possible nonconvexity of $\mathcal{F}$.

1.1.1 Example. Consider the bilevel program in $\Re^2$:

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & x \geq 0, \\ \text{and} & y \in \operatorname{argmin}\{y \,:\, y \in C(x)\}, \end{array}$$

where

$$C(x) \equiv \{y \in \Re_+ \,:\; x + 2y \geq 10,\; x - 2y \leq 6,\; 2x - y \leq 21,\; x + 2y \leq 38,\; -x + 2y \leq 18\}.$$

For $x \geq 0$, we have $C(x) \neq \emptyset$ if and only if $x \leq 16$. Since each inner problem is a linear program, we can solve for the optimal y for each given $x \in [0,16]$, obtaining

$$\mathcal{S}(x) = \begin{cases} 5 - x/2 & \text{if } x \in [0,8] \\ -3 + x/2 & \text{if } x \in [8,12] \\ -21 + 2x & \text{if } x \in [12,16]. \end{cases}$$

The feasible region $\mathcal{F}$, which is equal to

$$\{(x, 5 - x/2) \, : \, x \in [0, 8]\} \cup \{(x, -3 + x/2) \, : \, x \in [8, 12]\} \cup$$
$$\{(x, -21 + 2x) \, : \, x \in [12, 16]\}$$

is the union of three noncollinear line segments in the plane $\Re^2$. Thus $\mathcal{F}$ is nonconvex.

In the above example, the set $\mathcal{S}(x)$ is a singleton for each x. However, this solution function also illustrates another difficulty with the MPEC in general: namely, the function $\mathcal{S}(x)$ (in the single-valued case) is in general not Fréchet differentiable. Thus nonsmoothness is also an intrinsic feature of an MPEC.

Though not convex, the region $\mathcal{F}$ in Example **1.1.1** is at least a closed connected set. Next we give an example to show that this set $\mathcal{F}$ could be disconnected and not closed.

1.1.2 Example. Consider the bilevel program in $\Re^2$:

$$\begin{array}{ll} \text{minimize} & f(x, y) \\ \text{subject to} & |x| \leq 1, \\ \text{and} & \quad y \in \operatorname{argmin}\{y \, : \, |y| \leq 1, \, xy \leq 0\}. \end{array}$$

In the notation of the MPEC (1), we have $F(x, y) = 1$,

$$Z \;=\; \{(x, y) \in \Re^2 \, : \, |x| \leq 1\},$$

and

$$C(x) \;=\; \{y \in \Re \, : \, |y| \leq 1, \, xy \leq 0\}$$

is convex for each fixed x. It is not difficult to verify that

$$\mathcal{S}(x) \;=\; \begin{cases} \{-1\} & \text{if } x \in [0, 1] \\ \{0\} & \text{if } x \in [-1, 0). \end{cases}$$

Thus we have

$$\mathcal{F} \;=\; \{(x, -1) \, : \, x \in [0, 1]\} \cup \{(x, 0) \, : \, x \in [-1, 0)\}.$$

Clearly $\mathcal{F}$ is not closed.

The lack of closedness of $\mathcal{F}$ renders the MPEC more or less intractable. Subsequently, we shall impose some mild assumptions on the inner VIs that will allow us to circumvent this difficulty and focus on the case where $\mathcal{F}$ is indeed closed. Under these assumptions (that are satisfied by Example **1.1.1** but not by **1.1.2**), the feasible region $\mathcal{F}$ can be shown to be the union of finitely many closed sets. This structure of $\mathcal{F}$ brings out a combinatorial nature of the MPEC that adds to the difficulty of this problem. Indeed, the number of sets that constitute $\mathcal{F}$ could in general be large; they are the result of a complementarity condition implicit within the VIs.

Another difficulty with the MPEC is the multi-valued nature of the solution function $\mathcal{S}(x)$. This is illustrated by the following example.

1.1.3 Example. Consider the VI $(F(x,\cdot), C(x))$, where for all $(x,y) \in \Re^2$,

$$F(x,y) \equiv -y, \quad C(x) \equiv \{y \in \Re \,:\, |y| \le 1\}.$$

It can be verified that for all $x \in \Re$, $\mathcal{S}(x) = \{1,-1\}$ which is a discrete set.

A bilevel linear program is a special case of the MPEC in which f is a linear function in (x,y), Z is a polyhedron, F is a constant, and $C(x)$ is a polyhedron of special type:

$$\begin{array}{ll} \text{minimize} & c^T x + d^T y \\ \text{subject to} & A_1 x + A_2 y \ge a, \\ \text{and} & y \in \operatorname{argmin}\{q^T y \,:\, B_1 x + B_2 y \ge b\}, \end{array} \tag{7}$$

where the vectors a, b, c, d and matrices A_1, A_2, B_1, B_2 are of appropriate dimensions. In the notation of (1.1.1), we have

$$\begin{array}{llll} f(x,y) \equiv c^T x + d^T y, & Z \equiv \{(x,y) \in \Re^{n+m} \,:\, A_1 x + A_2 y \ge a\}, \\ F(x,y) \equiv q, & C(x) \equiv \{y \in \Re^m \,:\, B_1 x + B_2 y \ge b\}. \end{array}$$

It has been shown [122, 106] that the bilevel linear program belongs to the class of strongly NP-hard problems. (See [92] for an introduction to the theory of computational complexity and the definitions for various complexity classes of problems, such as that of P, NP, and strong NP.) This implies that there can be no fully polynomial approximation scheme for solving (7) unless the classes P and NP are equal. (Roughly speaking, a fully polynomial approximation scheme is an algorithm for computing an "ε-optimal" solution to a given problem with running time which is a polynomial in terms of the problem size and $1/\varepsilon$.) In spite of its hardness, the bilevel linear program has been well researched and there are many algorithms of the enumerative, branch-and-bound, exact-penalty, decomposition type for solving this problem [23, 24, 25, 106, 129, 279, 284].

The intrinsic difficulties of MPEC are unfortunate since this optimization problem has a wide range of applications in engineering and economics (the next section outlines several of these applications). Partly due to these difficulties, many studies of bilevel programs in the past have not been based on very sound principles and are full of loose arguments and heuristic approaches. Added to the complication is the fact that some of the early results reported in the literature are in fact incorrect. To illustrate, the reference [50] gave a counterexample to demonstrate that the necessary optimality conditions for the bilevel programming problem obtained in [18] were not correct; the reference [20] gave examples to show that several known methods claimed by their authors to always yield a globally optimal solution of a bilevel linear program were flawed. MPEC, being defined formally only recently [108], deserves to be given a comprehensive investigation and put on a solid, rigorous ground. The present monograph is written with this as its main objective.

1.2 Source Problems

Although the origin of the MPEC can be traced to the economic notion of a Stackelberg game [265], in the operations research literature mathematical programming problems with optimization constraints, i.e., bilevel programs, were introduced in a series of papers by Bracken and McGill [34, 36, 37, 35, 38]. Applications of these programs to military defense and production and marketing decision making in a competitive environment were also discussed in these references. As we shall see, the Bracken-McGill bilevel programs are considerably easier than the general MPEC. In the Ph.D. thesis [61], de Silva discussed the application of "an implicitly defined optimization model" to U.S. crude oil production. These and other early applications of MPEC are mostly concerned with bilevel programs where the inner problems are optimization problems. Along with the advance of the theory and methods for variational inequalities and complementarity problems [109] comes the gradual broadening of MPEC's applications to equilibrium modeling. The term "mathematical programs with equilibrium constraints" was coined in [108].

The volume [4] contains a number of interesting articles describing the diverse applications of hierarchical optimization in engineering and economics. In the next few subsections, we discuss some selective applications of the MPEC.

The Bracken-McGill bilevel programs

For historical reasons, we begin our discussion of the applications of MPEC with the earliest models proposed by Bracken and McGill. Their first few papers addressed bilevel programming models of minimum-cost weapon mix and other defense problems; in what follows, we present their optimal production and marketing decision making model published in [38].

Consider a firm which produces several products labeled $i = 1, \ldots, m$ using a number of different resources labeled $j = 1, \ldots, n$. The firm wishes to maximize profit subject to resource and market share constraints. The products are manufactured within resource availabilities, and certain minimum market shares must be maintained in the face of competition from other firms.

We introduce some notation. Let x_i and y_i denote, respectively, the firm's production and marketing level of the i-th product; let $x \equiv (x_i)$ and $y \equiv (y_i)$ be the n-vectors of production and marketing levels respectively. The real-valued function $g_i(x, y, u, v)$ expresses the firm's market share for product i given the values of x, y, u and v, where u and v are vectors denoting the competitors' production and marketing levels of all the products. The resource utilization function $h_j(x, y)$ specifies the amount of resource j that is required for production level x and marketing level y. The minimum fraction of market share for product i that is required by the firm is denoted a_i, and the total amount of resource j available to the firm is denoted b_j. Finally, let W denote the set of all feasible production and marketing levels for the competitors.

For a given level (x, y) of production and marketing, the firm's minimum market share function for the i-th product in the face of competition is given by

$$\sigma_i(x, y) \equiv \min\{g_i(x, y, u, v) \,:\, (u, v) \in W\}.$$

This function σ_i is in general not Fréchet differentiable, regardless of how smooth g_i is. The firm's optimal production and marketing strategy can be obtained by solving the following optimization problem in the variables (x, y):

$$\begin{array}{lll} \text{maximize} & f(x,y) & \\ \text{subject to} & h_j(x,y) \leq b_j, & j = 1, \ldots, n, \\ \text{and} & \sigma_i(x,y) \geq a_i, & i = 1, \ldots, m. \end{array} \tag{1}$$

In other words, the firm chooses a strategy (x, y) to maximize its total profit, subject to the minimum specified level of market share a_i for each product i, and to the resource limitations. In this formulation, the firm behaves in a rather conservative manner: it takes into account the worst-case scenario on the part of the competitors in order to ensure its desired market share. This approach is related to the weak Stackelberg game which will be described shortly.

The above problem can be put into the form of a bilevel program in the variables x, y, and $\{(u^i, v^i)\}_{i=1}^m$, where the new (lower-level) variables (u^i, v^i) are minimizers of the function $\sigma_i(x, y)$. More precisely, the problem (1) is equivalent to

$$\begin{array}{lll} \text{maximize} & f(x,y) & \\ \text{subject to} & h_j(x,y) \leq b_j, & j = 1, \ldots, n, \\ & g_i(x,y,u^i,v^i) \geq a_i, & i = 1, \ldots, m; \\ \text{and} & (u^i, v^i) \in \operatorname{argmin}\{g_i(x,y,u,v) \,:\, (u,v) \in W\}. & \end{array} \tag{2}$$

Notice that the lower-level variables $\{(u^i, v^i)\}_{i=1}^m$ do not appear in the upper-level objective function f. Moreover, under the assumption that $g_i(x, y, u^i, v^i)$ is concave in (x, y) for fixed (u^i, v^i) and W is a convex set, the function $\sigma_i(x, y)$ is concave in (x, y); if in addition $f(x, y)$ and h_j are concave in (x, y), then (1) is a concave maximization program. Thus the bilevel program (2) is considerably easier than the general MPEC (1.1.1) which is not expected to possess much convexity or concavity property.

Stackelberg game

The MPEC is intimately related to the so-called leader-follower (or Stackelberg) game [265, 13]. This game problem has been studied extensively by economists and has found wide application in such areas as oligopolistic market analysis [210, 263], optimal product design [48], quality control in services [6], and pricing of electric transmission [114]. The usage of the terms "leader" and "follower" in our introduction of the MPEC was derived from this Stackelberg game problem; see Section **1.1**.

The Stackelberg game can be considered an extension of the renowned Nash game [207]. In the Nash game, there are a number of (say, M) players each of whom has a strategy set $Y_i \subseteq \Re^{m_i}$. The objective of player i is to minimize its economic cost $\theta_i(y_i, y_{\neq i}^{\text{given}})$ by selecting a strategy $y_i \in Y_i$ given that the other players have chosen their strategies $y_{\neq i}^{\text{given}}$, where $y_{\neq i}^{\text{given}}$ denotes the vector $(y_j^{\text{given}} \; : \; j \neq i)$. In other words, each player observes the actions of the remaining players and then reacts optimally, assuming that the other players' strategies remain unchanged. A strategy combination $y^* \in \prod_{j=1}^{m} Y_j$ is called a Nash equilibrium if no player has an incentive to deviate from his strategy y_i^* in the sense that

$$y_i^* \in \operatorname{argmin}\{\theta_i(y_i, y_{\neq i}^*) \; : \; y_i \in Y_i\}, \qquad \forall i.$$

It should be noted that the players in the Nash game are in a sense homogeneous since each of them has access to the same information regarding the other players' strategies and the strategy chosen is only dependent on this information.

In contrast, the Stackelberg game has a distinctive player (called the leader) who can *anticipate* the (re)actions of the remaining players (called followers) and use this knowledge in selecting his optimal strategy (see [242]). Specifically, the leader chooses a strategy from the strategy set $X \subseteq \Re^n$, while each follower (say i) has, corresponding to each of the leader's strategies $x \in X$, a strategy set $Y_i(x) \subseteq \Re^{m_i}$ that is closed and convex and a cost function $\theta_i(x, \cdot) : \prod_{j=1}^{M} \Re^{m_j} \to \Re$, where M is the number of followers in the Stackelberg game. Note that each follower's strategy is dependent on the particular strategy x of the leader and this follower's cost function is dependent on both the leader's and all followers' strategies. Let $p \equiv \sum_{i=1}^{M} m_i$. We assume that for any fixed but arbitrary $x^{\text{given}} \in X$ and $y_{\neq i}^{\text{given}} \equiv (y_j^{\text{given}} : j \neq i)$, the function

$$\theta_i(x^{\text{given}}, y_i, y_{\neq i}^{\text{given}}) \tag{3}$$

is convex and continuous differentiable in the variable $y_i \in Y_i(x^{\text{given}})$.

Collectively, the followers behave according to the Nash noncooperative principle described above. That is to say, they will choose, for each $x \in X$, a joint response vector

$$y^{\text{opt}} \equiv (y_i^{\text{opt}})_{i=1}^M \in \prod_{i=1}^M Y_i(x)$$

such that for each $i = 1, \ldots, M$

$$y_i^{\text{opt}} \in \operatorname{argmin}\{\theta_i(x, y_i, y_{\neq i}^{\text{opt}}) \,:\, y_i \in Y_i(x)\}. \tag{4}$$

By the convexity of the payoff functions (3) and the sets $Y_i(x)$, it is easy to show that (4) holds for all $i = 1, \ldots, M$ if and only if the vector y^{opt} is in $\text{SOL}(F(x,\cdot), C(x))$ where for $y \in \Re^p$, $F(x,y) \equiv (F_i(x,y))_{i=1}^M$ with

$$F_i(x,y) \equiv \nabla_{y_i}\theta_i(x,y), \quad i = 1, \ldots, M,$$

and

$$C(x) \equiv \prod_{i=1}^M Y_i(x).$$

Let $f : \Re^{n+p} \to \Re$ be the leader's cost function which depends on both his own and the followers' strategies. The Stackelberg game problem is to determine a vector $(x,y) \in \Re^{n+p}$ in order to

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & x \in X, \\ \text{and} & y \in \mathcal{S}(x). \end{array} \tag{5}$$

This is an MPEC.

As we can see, in the Stackelberg game the leader is more "powerful" than the followers in the sense that the leader is allowed to anticipate the reactions of the followers and select his strategy accordingly. Thus, the players of a Stackelberg game are no longer homogeneous as in the case of the Nash game. Recall that in the Nash game there is no distinction among the players since they can only observe but not anticipate the (re)actions of the other players. If the leader loses the advantage of anticipating the other players' reactions then the Stackelberg game reduces to the standard Nash game. The loss of leader's privilege will usually result in an increase in the leader's optimal cost and a decrease in the followers' optimal costs; see [94] for an example. Similarly, if the leader's ability of anticipation does not affect his objective function value (e.g., $f(x, y) \equiv 0$), then the Stackelberg game also reduces to a Nash game.

Refinements and variations of the above Stackelberg model have been proposed and studied extensively by Jacqueline Morgan and her collaborators; see [157, 158, 165, 179]. In the terminology used in these references, our model is a *strong* or *optimistic* Stackelberg game. To understand this terminology, note that the problem (5) is equivalent to

$$\begin{array}{ll} \text{minimize} & f_{\text{strong}}(x) \equiv \inf_{y \in \mathcal{S}(x)} f(x, y) \\ \text{subject to} & x \in X. \end{array} \tag{6}$$

Thus the leader is optimistic in the sense that he assumes that the followers will act most favorably to his well-being by choosing, for each of his announced strategies $x \in X$, a reaction $y \in \mathcal{S}(x)$ among the rational reactions that will contribute to the minimization of the cost function f. This situation is to be contrasted with the *weak* or *pessimistic* Stackelberg game in which the leader assumes that the followers will choose their reactions from the rational set that will be least favorable to him; thus the leader will act conservatively to guard against the worst outcome. Mathematically, the leader will solve the following MPEC:

$$\begin{array}{ll} \text{minimize} & f_{\text{weak}}(x) \equiv \sup_{y \in \mathcal{S}(x)} f(x, y) \\ \text{subject to} & x \in X. \end{array}$$

When the reaction set $\mathcal{S}(x)$ is a singleton for each $x \in X$, there is no distinction between the above two situations. Nevertheless for Stackelberg games with a multivalued reaction map, the weak and strong versions are not necessarily equivalent. In [294], the terminology of "cooperative" and "noncooperative" has been used to mean "strong" and "weak" respectively. However, the former is confusing because a cooperative game normally has a somewhat different (and well established) meaning in game theory [216].

An application of the Stackelberg game model (5) in conjunction with the Cournot production model was discussed in [210, 263, 275]. The paper [48] proposes a Stackelberg game model for new product pricing and positioning in the face of price competition. Other studies on Stackelberg problems from an optimization point of view are [2, 5, 212, 213]. A numerical approach for computing a Stackelberg-Cournot-Nash equilibrium via nondifferentiable optimization is proposed in [215]. Incidentally, the bilevel programs introduced by Bracken and McGill, including the one discussed in the previous subsection, were developed to model a competitive environment similar to the leader-follower game discussed above. The difference is that in the Bracken-McGill model, there was no specification of the followers' (the competitors') behavior; also the way the leader (the firm) handled the followers' responses was different in the two approaches.

Misclassification minimization

Given two point sets $\mathcal{A}$ and $\mathcal{B}$ and a hyperplane H in an n-dimensional real Euclidean space, we nominate one side of H (a closed halfspace) as the A-side and the other as the B-side. We say that H *(linearly) separates* the two sets if the A-side can be chosen to contain all the points in $\mathcal{A}$ and the B-side all the points in $\mathcal{B}$. Such a separation is *strict* if no points of $\mathcal{A} \cup \mathcal{B}$ lie on H. When the two sets $\mathcal{A}$ and $\mathcal{B}$ are not linearly separated by a hyperplane H, then for any nomination of the A-side and B-side, there is some point in $\mathcal{A} \cup \mathcal{B}$ that lies in the wrong side of H (and not on H). Each of these misplaced points is called a *misclassification*. The problem of finding a hyperplane which separates $\mathcal{A}$ and $\mathcal{B}$ with a minimum number of misclassifications is of fundamental importance to the area of machine learning, pattern recognition, and artificial intelligence.

Mangasarian [187] has formulated the above misclassification problem as a bilevel linear program. Below is a brief description of this formulation. Suppose that the two sets of points $\mathcal{A}$, $\mathcal{B}$ are represented by the rows of two matrices A $(m \times n)$ and B $(k \times n)$ respectively, and suppose that the hyperplane H is given by

$$w^T x = \theta,$$

where $w \in \Re^n$ is the normal vector of the hyperplane and θ is a scalar; w and θ are to be determined. Clearly, the plane H separates the two sets $\mathcal{A}$ and $\mathcal{B}$ strictly if and only if

$$Aw > e\theta, \quad \text{and} \quad Bw < e\theta,$$

where e is a vector of ones of appropriate dimensions. By rescaling, the