THE UNIVERSITY OF MICHIGAN
COLLEGE OF ENGINEERING
Information and Control Engineering Program

COMPUTATION OF OPTIMAL CONTROLS BY QUADRATIC PROGRAMMING ON CONVEX REACHABLE SETS

Robert Ortha Barr, Jr.

ORA Projects 06999 and 08180

Supported by:
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
OFFICE OF AEROSPACE RESEARCH
ARLINGTON, VIRGINIA
Grant No. AF-AFOSR-814-65
and AF-AFOSR-814-66

Administered through:
OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

September 1966

UMR Ø225

This report was also a dissertation submitted in
partial fulfillment of the requirements for the
degree of Doctor of Philosophy in
The University of Michigan
1966

Abstract

COMPUTATION OF OPTIMAL CONTROLS BY QUADRATIC PROGRAMMING ON CONVEX REACHABLE SETS

by Robert Ortha Barr, Jr.

This thesis develops iterative procedures applicable to a wide variety of optimal control problems. The computational algorithms are quite different from those which have been described previously and apply under more general conditions. Convergence of the procedures is proved and numerical results for example problems are given.

Essential to all the procedures is a method for solving the basic problem BP: given K a compact, convex set in E^n; find a point $z^* \in K$ such that $|z^*|^2 = \min_{z \in K} |z|^2$ ($|\cdot|$ denotes Euclidean norm). The constraint set K in this quadratic programming problem, in contrast to the quadratic programming problems usually treated in the literature, need not be specified by some set of functional inequalities. It is required only that there be a known contact function of K, i.e., a function $s(\cdot)$ from E^n to K such that $y \cdot s(y) = \max_{z \in K} y \cdot z$ for $y \neq 0$. The method given by E. G. Gilbert for solving BP using only a contact function of K is extended so that on each iteration a quadratic minimization problem is solved on a convex polyhedron instead of on a line segment. As with Gilbert's method, computable error bounds are available. Techniques for readily solving the minimization problem on a convex polyhedron are discussed and extensive computational results for a rather general example using both Gilbert's method and its extension are presented. These results indicate that the extended procedure has a much improved rate of convergence. Furthermore, it is proved that if K is a convex polyhedron and the range of $s(y)$, $y \in E^n$, is a finite set of points, then the extended procedure exhibits finite convergence.

Certain optimal control problems are related directly to BP. However, it is possible to state an abstract problem which has application to a much broader class of optimal control problems. This problem, called the general problem GP, is: given a compact interval $\Omega = [0, \hat{\omega}]$, $\hat{\omega} > 0$, a family of sets $K(\omega)$ in E^n which are compact, convex, and continuous on Ω, and the fact that there exists $\omega^* \in \Omega$ such that $0 \notin K(\omega)$, $0 \leq \omega < \omega^*$, and $0 \in K(\omega^*)$; find ω^*. An iterative procedure for solving GP, which on each iteration involves the minimization problem BP, is described and shown to converge. Numerical results for this procedure applied to a minimum fuel example are given.

Optimal control problems are considered in which there are: i) a system differential equation of the form

$$\dot{x}(t) = A(t)\,x(t) + f(u(t), t), \quad x(0),$$

where x is the m-dimensional state vector, $x(0)$ is the initial state, $u(\cdot)$ is an r-dimensional control vector, admissible if measurable with range in a compact set U, the matrix function $A(t)$ and vector function $f(u, t)$ are continuous in their arguments; ii) compact, convex target sets $W(t)$ in E^m which are continuous in t; iii) a cost functional

$$J_t(u) = \int_0^t [a(\sigma) \cdot x_u(\sigma) + f^0(u(\sigma), \sigma)]\, d\sigma + h^0(x_u(t)),$$

where $x_u(t)$ is the solution of the system equation corresponding to the control $u(\cdot)$, $h^0(\cdot)$ is a convex function from E^m to E^1, the vector function $a(t)$ and scalar function $f^0(u, t)$ are continuous in their arguments; iv) an optimization objective corresponding to fixed or free terminal time. For a fixed terminal time $T > 0$ the objective is: find an admissible control $u^*(\cdot)$ such that $x_{u^*}(T) \in W(T)$ and the cost for $u^*(\cdot)$ at T $\leq$ the cost for $u(\cdot)$ at T where $u(\cdot)$ is any admissible control for which $x_u(T) \in W(T)$; if the terminal time is free: find an admissible control $u^*(\cdot)$ and an optimal time t^* such that $x_{u^*}(t^*) \in W(t^*)$ and the cost for

$u^*(\cdot)$ at $t^* \leq$ the cost for $\tilde{u}(\cdot)$ at $\tilde{t}$ where $\tilde{u}(\cdot)$ is any admissible control for which $x_{\tilde{u}}(\tilde{t}) \in W(\tilde{t})$, any $\tilde{t}$. The iterative procedures apply to many problems of this class and to other optimal control problems as well.

The set of all solutions of the system equation at a particular time t which are generated by admissible controls is the reachable set. This set is (1) compact, (2) convex, (3) continuous in t, and (4) has a contact function which can be evaluated. These four properties are the essential features which permit application of the iterative procedures to optimal control problems.

Acknowledgements

The author wishes to express his gratitude to each committee member, especially to Dr. E. G. Gilbert who provided thoughtful criticisms, encouragement, and suggestions throughout the course of this research. The helpful comments of Dr. P. Falb are also gratefully acknowledged.

The author is indebted to Mr. John A. Parsons for his excellent work in typing the manuscript.

Sponsorship for the research was provided by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR grant numbers AF-AFOSR-814-65 and AF-AFOSR-814-66.

Table of Contents

	Page
Acknowledgements	ii
List of Tables	v
List of Figures	ix
Chapter 1 Introduction	1
Chapter 2 The Basic Iterative Procedure BIP	4
2.1 Preliminary Comments on Notation	4
2.2 Contact Function; the Basic Problem BP	5
2.3 The Basic Iterative Procedure BIP	6
2.4 Convergence Theorem for BIP	9
2.5 Nature of Convergence; Examples	13
2.6 Numerical Results for BIP	18
Chapter 3 The Improved Iterative Procedure IIP	30
3.1 Convex Sets and Polyhedra	30
3.2 The Subproblem 1	34
3.3 The Improved Iterative Procedure IIP	35
3.4 Convergence Theorem for IIP	37
3.5 Selection Rules for IIP	39
3.6 Solution of Subproblem 1	48
3.7 A Finite Convergence Theorem	52
3.8 Numerical Results for IIP	55
Chapter 4 The General Iterative Procedure GIP	80
4.1 The General Problem GP	80
4.2 The General Iterative Procedure GIP	82
4.3 Convergence Theorem for GIP	86
Chapter 5 Application of BIP, IIP, and GIP to Optimal Control Problems	90
5.1 Some Optimal Control Problems Solvable by GIP	90

Page

5.2 The Reachable Set; Determination of a Contact Function 93

5.3 Solution of Problems 1 and 2 by BIP or IIP Alone 99

5.4 The Sets $K(\omega) = X(\omega) - Y(\omega)$ 102

5.5 Solution of Problems 1, 2, and 3 by GIP 103

5.6 Solution of Problems 4, 5, and 6 by GIP 105

5.7 Numerical Results for GIP Applied to a Minimum-Fuel Example 108

List of Symbols 128

Bibliography 133

List of Tables

Table		Page
2.6.1	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 2$, $z_0 = (6, 2)$, BIP.	23
2.6.2	Number of iterations to satisfy $\lvert z_k - z^* \rvert \le \epsilon$; $n = 2$, $z_0 = (6, 2)$, BIP.	23
2.6.3	Number of iterations to satisfy $\lvert z^* \rvert - \lvert z_k \rvert \gamma(z_k) \le \epsilon$; $n = 2$, $z_0 = (6, 2)$, BIP.	24
2.6.4	Number of iterations to satisfy $\lvert s(-z_k) - z^* \rvert \le \epsilon$; $n = 2$, $z_0 = (6, 2)$, BIP	24
2.6.5	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 2, \lambda_2 = 100$, BIP.	25
2.6.6	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 3$, $z_0 = (6, 2, 2)$, BIP.	28
2.6.7	Number of iterations to satisfy $\lvert z_k - z^* \rvert \le \epsilon$; $n = 3$, $z_0 = (6, 2, 2)$, BIP.	28
2.6.8	Number of iterations to satisfy $\lvert z^* \rvert - \lvert z_k \rvert \gamma(z_k) \le \epsilon$; $n = 3$, $z_0 = (6, 2, 2)$, BIP.	29
2.6.9	Number of iterations to satisfy $\lvert s(-z_k) - z^* \rvert \le \epsilon$; $n = 3$, $z_0 = (6, 2, 2)$, BIP.	29
3.8.1	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 2$, $p = 2$, $z_0 = (6, 2)$, IIP Selection Rule A.	61
3.8.2	Number of iterations to satisfy $\lvert z_k - z^* \rvert \le \epsilon$; $n = 2$, $p = 2$, $z_0 = (6, 2)$, IIP Selection Rule A.	61
3.8.3	Number of iterations to satisfy $\lvert z^* \rvert - \lvert z_k \rvert \gamma(z_k) \le \epsilon$; $n = 2$, $p = 2$, $z_0 = (6, 2)$, IIP Selection Rule A.	62

Table		Page
3.8.4	Number of iterations to satisfy $\lvert s(-z_k) - z^* \rvert \le \epsilon$; $n = 2$, $p = 2$, $\mathbf{z_0} = (6, 2)$, IIP Selection Rule A.	62
3.8.5	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 2$, $p = 2$, $\lambda_2 = 100$, IIP Selection Rule A.	63
3.8.6	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.	66
3.8.7	Number of iterations to satisfy $\lvert z_k - z^* \rvert \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.	66
3.8.8	Number of iterations to satisfy $\lvert z^* \rvert - \lvert z_k \rvert \gamma(z_k) \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.	67
3.8.9	Number of iterations to satisfy $\lvert s(-z_k) - z^* \rvert \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.	67
3.8.10	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z_k \rvert \gamma(z_k) \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.	68
3.8.11	Number of iterations to satisfy $\sqrt{1 - \gamma(z_k)}\, \lvert z_k \rvert \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.	68
3.8.12	Number of iterations to satisfy $\lvert s(-\mathbf{z}_k) \rvert - \lvert z^* \rvert \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.	69
3.8.13	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.	69
3.8.14	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule B.	70
3.8.15	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 2$, $p = 1$, $z_0 = (6, 2)$, IIP Selection Rule A.	71
3.8.16	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 2$, $z_0 = (6, 2)$, IIP Selection Rule A and BIP.	71

Table		Page
3.8.17	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 3$, $z_0 = (5, 4, 2)$, $\lambda_2 = 1000$, $\lambda_3 = 100$, IIP Selection Rule A and BIP.	72
3.8.18	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 4$, $p = 4$, $z_0 = (6, 2, 2, 1)$, IIP Selection Rule A.	73
3.8.19	Number of iterations to satisfy $\lvert z_k - z^* \rvert \le \epsilon$; $n = 4$, $p = 4$, $z_0 = (6, 2, 2, 1)$, IIP Selection Rule A.	74
3.8.20	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 4$, $z_0 = (6, 1, 2, 2)$, $\lambda_2 = 1000$, $\lambda_3 = 500$, $\lambda_4 = 100$, IIP Selection Rule A.	75
3.8.21	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; $n = 5$, $p = 5$, $z_0 = (5, 3, 1, 1.8, 2.6)$, IIP Selection Rule A.	76
3.8.22	Number of iterations to satisfy $\lvert z_k - z^* \rvert \le \epsilon$; $n = 5$, $p = 5$, $z_0 = (5, 3, 1, 1.8, 2.6)$, IIP Selection Rule A.	76
3.8.23	Number of iterations to satisfy error criteria; $n = 6$, $p = 6$, $z_0 = (4, 3, 2.6, 2.6, 1.8, 1.8)$, IIP Selection Rule A.	77
3.8.24	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; IIP Selection Rule C.	78
3.8.25	Number of iterations to satisfy $\lvert z_k \rvert - \lvert z^* \rvert \le \epsilon$; IIP Selection Rule C.	79
5.7.1	The first j for which $\bar{\delta}_j \le \epsilon$ and $\bar{\rho}_j \le \epsilon$; $x(0) = (2, -1)$.	114
5.7.2	The first j for which $\bar{\delta}_j \le \epsilon$ and $\bar{\rho}_j \le \epsilon$; $x(0) = (2, 0)$.	114
5.7.3	The first j for which $\bar{\delta}_j \le \epsilon$ and $\bar{\rho}_j \le \epsilon$; $x(0) = (7, -3)$; IIP Selection Rule A with $p = 3$ used on each iteration of GIP.	

Table		Page
5.7.4	The first j for which $\bar{\delta}_j \leq \epsilon$ and $\bar{\rho}_j \leq \epsilon$; x(0) = (7.5, -3); IIP Selection Rule A with p = 3 used on each iteration of GIP.	115
5.7.5	The functions ω_i and $\ell(i)$ for x(0) = (2, -1); IIP Selection Rule A with p = 3 used on each iteration of GIP.	116
5.7.6	The functions ω_i and $\ell(i)$ for x(0) = (2, -1); BIP used on each iteration of GIP.	117
5.7.7	The functions ω_i and $\ell(i)$ for x(0) = (2, 0); IIP Selection Rule A with p = 3 used on each iteration of GIP.	118
5.7.8	The functions ω_i and $\ell(i)$ for x(0) = (2, 0); BIP used on each iteration of GIP.	119
5.7.9	The functions ω_i and $\ell(i)$ for x(0) = (7, -3); IIP Selection Rule A with p = 3 used on each iteration of GIP.	120
5.7.10	The functions ω_i and $\ell(i)$ for x(0) = (7.5, -3); IIP Selection Rule A with p = 3 used on each iteration of GIP.	121

List of Figures

Figure		Page
2.3.1	Geometric significance of $\beta(\cdot)$ and $\gamma(\cdot)$.	7
2.3.2	Geometry of BIP for $\delta = 1$.	9
2.6.1	$\|z_k\| - \|z^*\|$ for $n = 2$, $z_0 = (6, 2)$: A) $\lambda_2 = 10$, B) $\lambda_2 = 100$, C) $\lambda_2 = 200$, D) $\lambda_2 = 500$, E) $\lambda_2 = 1000$; BIP.	21
2.6.2	$\beta(z_k)$ for $n = 2$, $z_0 = (6, 2)$, $\lambda\beta = 100$; BIP.	22
2.6.3	Results for $n = 3$, $z_0 = (6, 2, 2)$, $\lambda_2 = 100$, $\lambda_3 = 10$: A) $\|z^*\| \max_{i \le k} \|z_i\| \gamma(z_i)$, B) $\|z_k\| - \|z^*\|$, C) $\|z_k - z^*\|$; BIP. For $k \le 14$, $\|z_k - z^*\| \cong \|z_k\| - \|z^*\|$.	26
2.6.4	Results for $n = 3$, $z_0 = (6, 2, 2)$, $\lambda_2 = 100$, $\lambda_3 = 10$: A) $\|z^*\| - \|z_k\| \gamma(z_k)$, B) $\|s(-z_k) - z^*\|$; BIP.	27
3.5.1	Example 3 of Section 2.5, $n = 2$, $\nu = 1$, $\lambda_2 = 1$.	40
3.5.2	Example 1 of Section 2.5, $\nu = 1$.	41
3.7.1	A $K \subset E^2$ which does not satisfy assumption i) of Comment 3.7.2.	54
3.8.1	$\|z_k\| - \|z^*\|$ for $n = 2$, $p = 2$, $z_0 = (6, 2)$: A) $\lambda_2 = 100$, B) $\lambda_2 = 1000$; IIP Selection Rule A.	59
3.8.2	Results for $n = 2$, $p = 2$, $z_0 = (6, 2)$, $\lambda_2 = 100$: A) $\|z^*\| - \max_{i \le k} \|z_i\| \gamma(z_i)$, B) $\|z_k - z^*\|$; IIP Selection Rule A.	60
3.8.3	Results for $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, $\lambda_2 = 100$, $\lambda_3 = 10$: A) $\|z\| - \max_{i \le k} \|z_i\| \gamma(z_i)$, B) $\|z_k\| - \|z^*\|$, C) $\|z_k - z^*\|$; IIP Selection Rule A. For $k \le 9$, $\|z_k - z^*\| \cong \|z_k\| - \|z^*\|$.	64

Figure		Page
3.8.4	Results for $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$ $\lambda_2 = 100$, $\lambda_3 = 10$: A) $\lvert z^*\rvert - \lvert z_k\rvert \gamma(z_k)$, B) $\lvert s(-z_k) - z^*\rvert$; IIP Selection Rule A.	65
4.2.1	Notation for the case $\rho(\omega) \geq \epsilon$ in the proof of Theorem 4.2.1.	84
4.3.1	A configuration for which $\lvert z_{i+1}\rvert > \lvert z_i\rvert$ in GIP.	88
5.7.1	Initial conditions and isofuel contours for the minimum-fuel example.	109
5.7.2	Geometry for minimum-fuel example.	112
5.7.3	$\bar{\delta}_j$ and $\bar{\rho}_j$ for $x(0) = (2, -1)$, $\theta = .4$; BIP used on each iteration of GIP.	122
5.7.4	$\bar{\delta}_j$ and $\bar{\rho}_j$ for $x(0) = (2, -1)$, $\theta = .4$; IIP Selection Rule A with $p = 3$ used on each iteration of GIP.	123
5.7.5	$\bar{\delta}_j$ and $\bar{\rho}_j$ for $x(0) = (2, 0)$, $\theta = .4$; BIP used on each iteration of GIP.	124
5.7.6	$\bar{\delta}_j$ and $\bar{\rho}_j$ for $x(0) = (2, 0)$, $\theta = .4$; IIP Selection Rule A with $p = 3$ used on each iteration of GIP.	125
5.7.7	$\bar{\phi}_j$ for $x(0) = (2, -1)$, $\theta = .4$: A) IIP Selection Rule A with $p = 3$ used on each iteration of GIP; B) BIP used on each iteration of GIP.	126
5.7.8	$\bar{\phi}_j$ for $x(0) = (2, 0)$, $\theta = .4$: A) IIP Selection Rule A with $p = 3$ used on each iteration of GIP; B) BIP used on each iteration of GIP.	127

CHAPTER 1

INTRODUCTION

Recent advances in engineering and science, especially in space technology, have stimulated much interest in optimal controls. The availability of modern computers makes it feasible to consider the actual computation of these controls. The central theme of this dissertation is the development of new and improved computational methods for a broad class of optimal control problems.

Some of the early work in this area concerned rather specialized low-order problems in which an explicit solution could be obtained. For example, Bushaw [B7] solved the second-order minimum time regulator by geometrical constructions in the phase plane. More recently, because of the need for solving complex problems, a great variety of general computational methods have appeared in the literature. The method of dynamic programming introduced by Bellman [B8] has application to many optimization problems. However, the large storage capacity and long computational times that are usually required limit the practical usefulness of the technique. Gradient methods in function space also have been proposed [e.g., B11, K1] and are quite commonly used, but are difficult to apply in problems with control constraints and usually exhibit slow convergence. The two-point boundary value problems which arise from the necessary conditions of Pontryagin [P2] or variational calculus provide the starting point for many other computational methods. For example, there are techniques [B9, B10, K2, P1] which consider certain mappings of the boundary values at the initial or final time. There are also methods which are based on the convexity of sets related to the reachable set of system states. Such convexity methods are the principal concern in this thesis.

Most of the convexity methods involve gradient-type minimization of scalar functions of n variables [e.g., B1, E1, F1, F2, N1, N3, N4]. These techniques possess inherent difficulties of step-size determination and exhibit slow convergence in the presence of poor problem conditioning. Furthermore, they require an assumption that the reachable sets be strictly convex. This assumption is not satisfied in certain important problems; even if it is satisfied, it may be difficult to verify. Another convexity method, which provided the motivation for this dissertation, is that due to Gilbert [G1]. Gilbert's method and the procedures developed here are not gradient-type techniques and do not require strict convexity. Other desirable features are: rapid monotone convergence and the availability of computable error bounds. It should be noted that Gilbert's investigations in this area began by proving convergence for an extension by Fancher [F3] of a procedure due to Ho [H3].

The outline of the dissertation is as follows. In Chapter 2 a basic quadratic programming problem BP is stated and Gilbert's algorithm for solving it is presented in detail. The results of many numerical computations with this algorithm are also given. Gilbert's method requires a quadratic minimization problem on a line segment to be solved on each iteration. Chapter 3 contains an extension of Gilbert's method in which quadratic minimization problems are solved on convex polyhedra. The theory is carefully developed and includes a convergence proof and a sufficient condition for convergence in a finite number of iterations. Extensive computational results for a rather general example are presented. These results indicate that the extended procedure for solving BP is a significant improvement over Gilbert's method. In Chapter 4 a general problem GP which has application to a wide variety of optimal control problems is formulated. The basic idea of the formulation is due to Fadden [F1], but the iterative procedure for solving GP and the proof of convergence are new. Finally, in Chapter 5 the iterative procedures are applied to optimal control problems of a broad class. Numerical results for a minimum fuel example are also given.

CHAPTER 2

THE BASIC ITERATIVE PROCEDURE BIP

Gilbert [Gl] has presented an iterative procedure for computing the minimum of a quadratic form on a compact convex set in finite dimensional space. This algorithm will be called the basic iterative procedure since it is fundamental to all the computing procedures developed in this thesis. The first five sections of this chapter are devoted to the theoretical details of Gilbert's method and except for minor changes are taken directly from his paper [Gl]. In Section 2.6 the results of the author's extensive computer experimentation with Gilbert's procedure are presented.

2.1 Preliminary Comments on Notation

Let $x = (x^1, x^2, \ldots, x^n)$, $y = (y^1, y^2, \ldots, y^n)$, $z = (z^1, z^2, \ldots, z^n)$ be vectors in n-dimensional Euclidean space E^n and let ω be a scalar. The following notation is employed: $x \cdot y = \sum_{i=1}^{n} x^i y^i$, the scalar product of x and y; $|x| = (x, x)^{\frac{1}{2}}$, the Euclidean norm of x, and $|x - y|$, the distance between points x and y; $L(x;y) = \{z : z = x + \omega(y - x), -\infty < \omega < \infty\}$, $x \neq y$, the line passing through x and y; $N(x;\omega) = \{z : |z - x| < \omega\}$, $\omega > 0$, the open sphere with center at x and radius ω; $\overline{N}(x;\omega) = \{z : |z - x| \leq \omega\}$, the corresponding closed sphere; $Q(x;y) = \{z : z \cdot y = x \cdot y\}$, $y \neq 0$, the hyperplane (dimension n - 1) through x with normal y; $Q^+(x;y) = \{z : z \cdot y < x \cdot y\}$, $y \neq 0$, the open half-space bounded by Q(x;y) with outward normal y; $Q^-(x;y) = \{z : z \cdot y \geq x \cdot y\}$, $y \neq 0$, the closed half-space bounded by Q(x;y) with inward normal y. In addition, if X and Y are arbitrary sets in E^n: ∂X denotes the boundary of X; X + Y is the set $\{z : z = x + y, x \in X, y \in Y\}$; X - Y is the set $\{z : z = x - y, x \in X, y \in Y\}$.

2.2 Contact Function; the Basic Problem BP

Consider a set $K \subset E^n$ which is compact and convex. The real-valued function $\eta(y) = \max_{z \in K} z \cdot y$ is called the support function of K. Since K is compact, $\eta(y)$ is defined for all y. Furthermore, it can be shown that $\eta(y)$ is a convex function on E^n, a result which implies that $\eta(y)$ is continuous on E^n [E2]. Let P(y), $y \neq 0$, be the hyperplane $\{x : x \cdot y = \eta(y)\}$. Since $z \cdot y \leq \eta(y)$ for all $z \in K$ and $P(y) \cap K$ is not empty, P(y) is the (unique) support hyperplane of K with outward normal y. For each $y \neq 0$ the set $S(y) = P(y) \cap K$ is called the contact set of K and its elements are called contact points of K. It follows that S(y) is not empty, $S(y) \subset \partial K$, $S(\omega y) = S(y)$ for $\omega > 0$. If for every $y \neq 0$, S(y) contains only a single point, K is strictly convex.

A function s(y) mapping E^n into E^n is defined to be a contact function of K if $s(y) \in S(y)$, $y \neq 0$, and $s(0) \in K$. From the preceding it may be concluded that: $s(\cdot)$ is bounded; $s(y) = s(\omega y)$, $\omega > 0$; $\eta(y) = s(y) \cdot y$. Furthermore, on the set $\{y : |y| > 0\}$ each of the following is true if and only if K is strictly convex: $s(\cdot)$ is uniquely determined, $s(\cdot)$ is continuous. The continuity result is proved in [N1].

If for every y there is a method for determining a point $x(y) \in K$ such that $x(y) \cdot y = \max_{z \in K} z \cdot y = \eta(y)$, then it is said that a contact function of K is available. Such a determination, which corresponds to the solution of a linear programming problem on K, is possible for the convex sets which arise in a variety of optimal control problems (see Chapter 5). This availability of a contact function is essential to all the computing procedures which follow. Consider now the basic problem:

BP Given: K a compact, convex set in E^n. Find: a point $z^* \in K$ such that $|z^*| = \min_{z \in K} |z|$.

Since K is compact and $|z|$ is a continuous function of z, a solution z^* exists. The following additional results hold:

THEOREM 2.2.1 (Solution Properties for BP). i) z^* is unique; ii) $|z^*| = 0$ if and only if $0 \in K$; iii) for $|z^*| > 0$, $z^* \in \partial K$; iv) for $|z^*| > 0$, $z = z^*$ if and only if $z \in P(-z) \cap K = S(-z)$.

Proof: Properties ii) and iii) are obvious. Property i) is proved by contradiction. Suppose z_1^* and z_2^* are distinct solutions. Then by convexity $\tilde{z} = \frac{1}{2}z_1^* + \frac{1}{2}z_2^* \in K$, which means $|\tilde{z}| \geq |z_1^*| = |z_2^*|$. But this implies $|\frac{1}{2}z_1^* + \frac{1}{2}z_2^*|^2 \geq \frac{1}{2}|z_1^*|^2 + \frac{1}{2}|z_2^*|^2$ which can be written $|z_1^* - z_2^*|^2 \leq 0$, an inequality which is only true for $z_1^* = z_2^*$. Consider iv). The condition $z \in P(-z) \cap K$ implies $z \in P(-z) = Q(z;z)$. But $Q(z;z)$ is the support hyperplane for the closed sphere $\overline{N}(0;|z|)$ whose outward normal is z and whose contact point is z. Therefore, $Q(z;z)$ is a (separating) support hyperplane for K and $\overline{N}(0;|z|)$. Thus $K \cap N(0;|z|)$ is empty. Since $z \in K \cap \overline{N}(0;|z|)$, this implies $z = z^*$. The steps of this argument may be reversed to obtain the converse result.

Since $|z|$ is the Euclidean norm of z, the above minimization problem is equivalent to finding $z^* \in K$ such that $|z^*|^2 = \min_{z \in K} |z|^2$. Thus BP is a quadratic programming problem. It differs from the quadratic programming problems which are frequently treated in the literature [e.g., A1, B2, H1, V1] in that the constraint set is not described by some set of known algebraic equations or inequalities. Information about the set K is obtainable only through a contact function $s(\cdot)$.

In actual computations it is generally not possible to obtain z^*. However, it is important to know if $z \in K$ is such that $|z| - |z^*| \leq \epsilon_1$ and $|z - z^*| \leq \epsilon_2$ where ϵ_1 and ϵ_2 are specified positive constants. Expressions for determining if these inequalities are satisfied are given as part of the iterative procedures which follow.

2.3 The Basic Iterative Procedure BIP

In this section Gilbert's algorithm for computing the solution to

BP is described.

As a first step, let $s(\cdot)$ be a specific contact function of K and consider:

$$\beta(z) = |z - s(-z)|^{-2} z \cdot (z - s(-z)), \qquad z - s(-z) \neq 0 \qquad (2.3.1)$$
$$= 0, \qquad z - s(-z) = 0$$

and

$$\gamma(z) = |z|^{-2} z \cdot s(-z), \quad |z| > 0, \qquad z \cdot s(-z) > 0 \qquad (2.3.2)$$
$$= 0, \qquad z = 0 \text{ or } |z| > 0, \qquad z \cdot s(-z) \leq 0.$$

Thus $\beta(\cdot)$ and $\gamma(\cdot)$ are functions which are defined on K. Figure 2.3.1 indicates their geometric significance: $x = z + \beta(z)(s(-z) - z)$ is the point on the line $L(z;s(-z))$ with minimum Euclidean length; $\gamma(z)z$ is either the point $L(0;z) \cap P(-z)$ or the origin, depending on whether or not $L(0;z) \cap P(-z)$ is on the line segment connecting z and the origin.

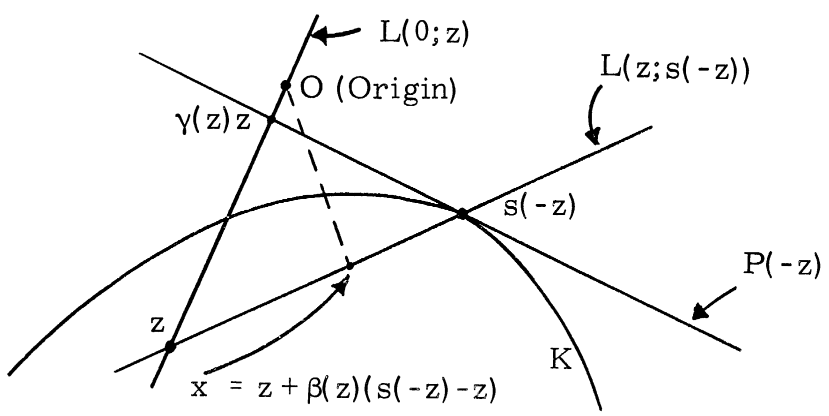


Figure 2.3.1 Geometric significance of $\beta(\cdot)$ and $\gamma(\cdot)$.

The functions $\beta(\cdot)$ and $\gamma(\cdot)$ have the following properties.

THEOREM 2.3.1 *Let* K *be the set described in* BP *and restrict* z *to* K. *Then*: i) $\beta(z) \geq 0$; ii) $\beta(z) = 0$ *if and only if* $z = z^*$; iii) $0 \leq \gamma(z) \leq 1$; iv) *if* $0 \in K$, $\gamma(z) \equiv 0$; v) *if* $0 \notin K$, $\gamma(z) = 1$ *if and only if*

$z = z^*$; vi) $\gamma(z)$ is continuous.

Proof: In this paragraph z always denotes a point in K. Later in this chapter (inequality (2.4.5)) it is shown that $0 \le z \cdot (z - s(-z))$. Hence, i) and iii) follow from (2.3.1) and (2.3.2). For the time being assume $|z^*| > 0$. The conditions $\beta(z) = 0$ and $\gamma(z) = 1$ both imply $z \cdot (-z) = s(-z) \cdot (-z) = \eta(-z)$ which requires $z \in P(-z)$. Since $z \in K$, part iv) of Theorem 2.2.1 yields $z = z^*$. Reversing these arguments completes the proof of ii) for $|z^*| > 0$ and of v). Now take $|z^*| = 0$. Inequality (2.4.4) then implies $s(-z) \cdot z \le 0$ which by (2.3.2) yields iv). If $\beta(z) = 0$ then it must follow from (2.3.1) that $s(-z) \cdot z = |z|^2$. Because of $s(-z) \cdot z \le 0$ this implies $z = 0 = z^*$. Since $z = z^* = 0$ also yields $\beta(z) = 0$, the proof of ii) is complete. For $|z| \ge |z^*| > 0$, the continuity of $\gamma(z)$ follows from (2.3.2) and the continuity of the support function $\eta(y) = s(y) \cdot y$. For $|z^*| = 0$, it is trivially true from iv).

It is of interest to note that $\beta(\cdot)$ may be discontinuous on K even though $s(\cdot)$ is continuous on K. See Example 3, Section 2.5.

For δ a fixed number, $0 < \delta \le 1$, and $z \in K$, define the closed interval

$$I(z) = [\min\{\delta\beta(z), 1\}, \ \min\{(2-\delta)\beta(z), 1\}] \ . \tag{2.3.3}$$

Then Gilbert's algorithm may be stated as follows:

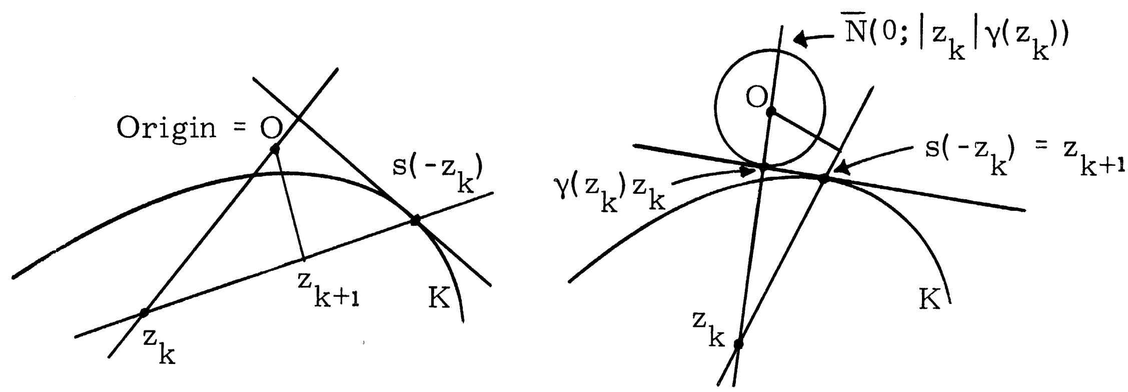
The Basic Iterative Procedure BIP Let $s(\cdot)$ be an arbitrary contact function of the set K specified in BP. Take $z_0 \in K$ and choose δ in the interval $0 < \delta \le 1$. Then a sequence of vectors $\{z_k\}$ in E^n is generated according to the rule

$$z_{k+1} = z_k + \alpha_k(s(-z_k) - z_k), \ k = 0, 1, 2, \ldots, \tag{2.3.4}$$

where scalars α_k are selected arbitrarily from $I(z_k)$.

For the case $\delta = 1$ the selection of $\alpha_k \in I(z_k)$ reduces to $\alpha_k = \text{sat}\ \beta(z_k)$, where sat ω is ω for $0 \le \omega \le 1$ and is 1 for $\omega > 1$. Figure

2.3.2 gives the geometric interpretation of BIP for this case.



$\beta(z_k) = \alpha_k < 1, \quad \gamma(z_k) = 0$ $\qquad$ $\beta(z_k) > 1, \quad \alpha_k = 1, \quad \gamma(z_k) > 0$

Figure 2.3.2 Geometry of BIP for $\delta = 1$.

If $\beta(z_k) > 0$ an improvement is obtained on the kth step, i.e., $|z_{k+1}| < |z_k|$; if $\beta(z_k) = 0$, $z_k = z^*$ and the iterative procedure is finite, i.e., the solution has been obtained in k steps. From Figure 2.3.2 it is also clear that $|z_k|\gamma(z_k) \le |z^*| \le |z_k|$. Thus on each step upper and lower bounds on $|z^*|$ may be computed. Notice that in applying the iterative procedure it is not necessary to know beforehand whether or not $0 \in K$. A more precise and complete statement of results is contained in the theorem of the next section.

2.4 Convergence Theorem for BIP

THEOREM 2.4.1 Consider the sequence $\{z_k\}$ generated by BIP. For $k \ge 0$ and $k \to \infty$: i) $z_k \in K$; ii) the sequence $\{|z_k|\}$ is decreasing ($|z_k| \ge |z_{k+1}|$), $|z_k| \to |z^*|$, and $|z_k| = |z_{k+1}|$ implies $z_k = z^*$; iii) $z_k \to z^*$; iv) $|z_k|\gamma(z_k) \le |z^*|$ and $|z_k|\gamma(z_k) \to |z^*|$; v) $|z_k - z^*| \le \sqrt{1 - \gamma(z_k)}\ |z_k|$ and $\sqrt{1 - \gamma(z_k)}\ |z_k| \to 0$; vi) $|s(-z_k) - z^*| \le |s(-z_k) - \gamma(z_k)z_k|$.

Since the bounds given in parts iv), v), and vi) are computable as the iterative process proceeds, they may be used to generate stopping criteria for the termination of the iterative process. Example problems show $\{|z_k|\gamma(z_k)\}$ is not necessarily increasing. Thus $|z_k| - \max_{i \le k} |z_i|\gamma(z_i)$ is more satisfactory as an upper bound for $|z_k| - |z^*|$ than $|z_k| - |z_k|\gamma(z_k)$. Since examples also show that $\{|z_k - z^*|\}$ and $\{|s(-z_k) - z^*|\}$ are not necessarily decreasing, it is not possible to improve similarly the bounds given in v) and vi).

Suppose $|z^*| > 0$ and $s(\cdot)$ is continuous in a neighborhood of $-z^*$ (the latter is certainly implied if K is strictly convex). Then it follows from the continuity of $\gamma(\cdot)$ and part iii) that the upper bound in part vi) converges to zero. Thus $\{s(-z_k)\}$ may be used as an approximating sequence, an approach which may be advantageous in some situations. In addition it is clear from part iv) that $|s(-z_k)| - |z^*| \le |s(-z_k)| - \max_{i \le k} |z_i|\gamma(z_i)$, where the right side converges to zero. Therefore meaningful stopping criteria are available.

Proof of Theorem 2.4.1 (due to Gilbert [Gl]): First some basic inequalities are stated. From $z^* \in P(-z^*)$, $0 \notin K$, and $s(-y) \in P(-y)$, $y \neq 0$, it follows by the definition of $P(\cdot)$ that

$$z^* \cdot z^* \le z \cdot z^*, \qquad z \in K, \quad 0 \notin K; \tag{2.4.1}$$

$$s(-y) \cdot y \le z \cdot y, \qquad z \in K, \quad y \in E^n. \tag{2.4.2}$$

These inequalities lead to:

$$|z^*|^2 \le s(-y) \cdot z^*, \qquad 0 \notin K, \ y \in E^n; \tag{2.4.3}$$

$$s(-y) \cdot y \le z^* \cdot y, \qquad y \in E^n; \tag{2.4.4}$$

$$s(-z) \cdot z \le |z|^2, \qquad z \in K; \tag{2.4.5}$$

$$|y - z^*|^2 + z^* \cdot (y - z^*) \le y \cdot (y - s(-y)), \qquad y \in E^n; \tag{2.4.6}$$

$$|z - z^*|^2 \le z \cdot (z - s(-z)), \quad z \in K, \qquad 0 \notin K. \tag{2.4.7}$$

Inequalities (2.4.3), (2.4.4), and (2.4.5) are deduced from (2.4.1) and (2.4.2 by obvious substitutions. From the identity $|y - z^*|^2 + z^* \cdot (y - z^*) + y \cdot (z^* - s(-y)) = y \cdot (y - s(-y))$ inequality (2.4.6) follows by (2.4.4). Inequality (2.4.7) follows from (2.4.6) by use of (2.4.1).

Part i) of the theorem depends on $\alpha_k \in I(z_k)$ which insures $0 \le \alpha_k \le 1$. Thus from (2.3.4), $s(-z_k) \in K$, and the convexity of K, $z_k \in K$ implies $z_{k+1} \in K$.

Consider now the inequalities in parts iv), v), and vi). From (2.4.4) and the Schwartz inequality $s(-y) \cdot y \le |y| \cdot |z^*|$. Thus iv) follows from (2.3.2). The proof of the inequalities in parts v) and vi) makes use of $z = z_k \in K$. For $s(-z) \cdot z > 0$, $z \cdot (z - s(-z)) = |z|^2(1 - \gamma(z))$ and from (2.4.7) the inequality in v) is true. Now consider $s(-z) \cdot z \le 0$, which corresponds to $\gamma(z) = 0$. For $z^* = 0$, $\gamma(z) = 0$ (Theorem 2.3.1) and v) holds as an equality; for $z^* \neq 0$, the inequality in v) follows from (2.4.1) which insures $|z - z^*|^2 = |z|^2 - 2z \cdot z^* + 2|z^*|^2 - |z^*|^2 \le |z|^2$. If $z^* = 0$ the inequality in vi) is trivially true. Consider now $z^* \neq 0$. If $s(-z) \cdot z \le 0$, part vi) reduces to $-2s(-z) \cdot z^* + |z^*|^2 \le 0$ which is true by (2.4.3). The following identity is easily verified $|s(-z) - z^*|^2 = |s(-z) - \gamma(z)z|^2 + |z|^{-2}(s(-z) \cdot z)^2 + |z^*|^2 - 2s(-z) \cdot z^*$. Assuming $s(-z) \cdot z > 0$ and using $s(-z) \cdot z \le |z| \cdot |z^*|$ yields $|z|^{-2}(s(-z) \cdot z)^2 \le |z^*|^2$. Thus $|s(-z) - z^*|^2 \le |s(-z) - \gamma(z)z|^2 + 2(|z^*|^2 - s(-z) \cdot z^*)$ and by (2.4.3) the inequality in vi) follows.

In order to complete the proof of the theorem, the function

$$\Gamma(z) = |z|^2 - |z^*|^2 = |z - z^*|^2 + 2z^* \cdot (z - z^*) \tag{2.4.8}$$

is introduced. For $0 \notin K$ inequality (2.4.1) gives

$$0 \le |z - z^*|^2 \le \Gamma(z), \quad z \in K, \tag{2.4.9}$$

a result which is obviously true for $0 \in K$. In the following paragraphs it will be shown that $\{\Gamma(z_k)\}$ is decreasing and $\Gamma(z_k) \to 0$. By (2.4.8) and (2.4.9) this proves the first two results in ii) and iii). From

(2.3.4) $|z_{k+1}| \le |z_k| + \alpha_k |s(-z_k) - z_k|$ so $|z_k| = |z_{k+1}|$ implies $\alpha_k |s(-z_k) - z_k| = 0$. Thus either $\alpha_k = 0$ or $z_k = s(-z_k)$, both of which yield $\beta(z_k) = 0$ by (2.3.3) and (2.3.1). Then part ii) of Theorem 2.3.1 implies $z_k = z^*$, which completes the proof of ii). The remaining results in parts iv) and v) follow from the known value of $\gamma(z^*)$, the continuity of $\gamma(\cdot)$, and part iii).

For simplicity let

$$\bar{\Delta}(z;\alpha) = \Gamma(z) - \Gamma(z + \alpha(s(-z) - z)) \tag{2.4.10}$$

and assume tacitly in what follows that $z \in K$. Then from (2.4.8)

$$\bar{\Delta}(z;\alpha) = 2\alpha(|z|^2 - s(-z) \cdot z) - \alpha^2 |z - s(-z)|^2 \tag{2.4.11}$$

Because the coefficient of α^2 is not positive, $\min_{\alpha \in I(z)} \bar{\Delta}(z;\alpha)$ is attained at the end points of $I(z)$. It is readily shown that $\bar{\Delta}(z;\delta\beta(z)) = \bar{\Delta}(z;(2-\delta)\beta(z))$. Thus from the definition of $I(z)$,

$$\begin{aligned} \min_{\alpha \in I(z)} \bar{\Delta}(z;\alpha) &= \bar{\Delta}(z;\delta\beta(z)) , \quad && \beta(z) \le \delta^{-1} \\ &= \bar{\Delta}(z;1) , && \beta(z) \ge \delta^{-1} . \end{aligned} \tag{2.4.12}$$

Equation (2.4.12) is now used to obtain a lower bound on $\bar{\Delta}(z;\alpha)$, $\alpha \in I(z)$. From (2.4.11) and (2.3.1) it follows that

$$\bar{\Delta}(z;\delta\beta(z)) = |z - s(-z)|^{-2} [z \cdot (z - s(-z))]^2 (2\delta - \delta^2). \tag{2.4.13}$$

Let

$$\mu = \max_{z_1, z_2 \in K} |z_1 - z_2| \tag{2.4.14}$$

denote the diameter of K and recall that $0 < \delta \le 1$. Then

$$\bar{\Delta}(z;\delta\beta(z)) \ge \mu^{-2} \delta [z \cdot (z - s(-z))]^2 \tag{2.4.15}$$

From (2.4.8) and (2.4.6)

$$\Gamma(z) \le 2|z - z^*|^2 + 2z^* \cdot (z - z^*) \le 2z \cdot (z - s(-z)) \tag{2.4.16}$$

(for $z^* = 0$ this may be sharpened to $\Gamma(z) \le z \cdot (z - s(-z))$). Thus

$$\bar{\Delta}(z;\delta\beta(z)) \geq \tfrac{1}{4}\mu^{-2}\delta\Gamma^2(z). \tag{2.4.17}$$

For $\beta(z) \geq 1$, $z \cdot (z - s(-z)) \geq |z - s(-z)|^2$ and consequently

$$\bar{\Delta}(z;1) = 2z \cdot (z - s(-z)) - |z - s(-z)|^2 \geq z \cdot (z - s(-z)).$$

Therefore (2.4.16) yields

$$\bar{\Delta}(z;1) \geq \tfrac{1}{2}\Gamma(z), \quad \beta(z) \geq 1. \tag{2.4.18}$$

Finally, utilizing (2.4.17) and (2.4.18) in (2.4.12) yields

$$\bar{\Delta}(z;\alpha) \geq \min\{\tfrac{1}{4}\mu^{-2}\delta\Gamma^2(z), \ \tfrac{1}{2}\Gamma(z)\} \tag{2.4.19}$$
$$\alpha \in I(z)$$

Letting $z = z_k$ in (2.4.19), using (2.3.4), and returning to (2.4.10), it is seen that

$$\Gamma(z_k) - \Gamma(z_{k+1}) \geq \min\{\tfrac{1}{4}\mu^{-2}\delta\Gamma^2(z_k), \ \tfrac{1}{2}\Gamma(z_k)\} \geq 0. \tag{2.4.20}$$

Therefore the sequence $\{\Gamma(z_k\}$ is decreasing and, since it is bounded from below by zero, has a limit point. Thus passing to the limit on the left side of (2.4.20) gives zero and therefore from the right side $\Gamma(z_k) \to 0$.

2.5 Nature of Convergence; Examples

This section gives some further results of Gilbert on the convergence of BIP. Theorem 2.5.1 establishes upper bounds on the elements of the sequences $\{|z_k|\}$ and $\{|z_k - z^*|\}$. Three example problems are analyzed to demonstrate still more fully the nature of convergence. Emphasis is on the case $0 \notin K$, since it is most important in applications.

THEOREM 2.5.1 *Let*

$$\theta_k = \theta_0(1 + \tfrac{1}{4}\mu^{-2}\delta\theta_0 k)^{-1}, \quad \theta_0 = |z_0|^2 - |z^*|^2 \tag{2.5.1}$$

and assume that $|z_0|^2 \leq |z^*|^2 + 2\mu^2\delta^{-1}$. *Then if* $\{z_k\}$ *is generated by*

BIP, <u>the following inequalities hold for</u> $k \geq 0$:

$$|z_k| \leq \sqrt{\theta_k + |z^*|^2} \; , \tag{2.5.2}$$

$$|z_k - z^*| \leq \sqrt{\theta_k} \; . \tag{2.5.3}$$

The assumption on $|z_0|$ is often met in practice. For example, it is easily shown that it must be satisfied if $|z^*| \leq \frac{1}{2}(2\delta^{-1} - 1)\mu$. In any case, z_0 may be interpreted as a suitable intermediate point in the iterative process, and inequalities (2.5.2) and (2.5.3) may be used to estimate the subsequent rate of convergence.

For $|z^*| > 0$ and $k \geq 1$ inequalities (2.5.2) and (2.5.3) imply

$$|z_k| - |z^*| < 2\mu^2 |z^*|^{-1} \delta^{-1} k^{-1} , \tag{2.5.4}$$

$$|z_k - z^*| < 2\mu\delta^{-\frac{1}{2}} k^{-\frac{1}{2}} , \tag{2.5.5}$$

results which conform closely to (2.5.2) and (2.5.3) for k sufficiently large. In Examples 1 and 2, which appear later in this section, it is demonstrated that within a constant multiplicative factor it is impossible to obtain bounds on $|z_k| - |z^*|$ and $|z_k - z^*|$ which approach zero more rapidly than those given in (2.5.4) and (2.5.5)

<u>Proof of Theorem 2.5.1</u>: Since $|z_0|^2 \leq |z^*|^2 + 2\mu^2\delta^{-1}$ it follows from the previous section that $\Gamma(z_k) \leq \Gamma(z_0) \leq 2\mu^2\delta^{-1}$, $k \geq 0$. From (2.4.20) this implies $\Gamma(z_{k+1}) \leq \Gamma(z_k) - \frac{1}{4}\mu^{-2}\delta\Gamma^2(z_k)$, $k \geq 0$. Since $1 - \frac{1}{4}\mu^{-2}\delta\Gamma \leq (1 + \frac{1}{4}\mu^{-2}\delta\Gamma)^{-1}$ for all $\Gamma \geq 0$, it is possible to write

$$\Gamma(z_{k+1}) \leq \Gamma(z_k)(1 + \tfrac{1}{4}\mu^{-2}\delta\Gamma(z_k))^{-1}, \quad k \geq 0. \tag{2.5.6}$$

But substitution shows that θ_k is the solution of

$$\theta_{k+1} = \theta_k(1 + \tfrac{1}{4}\mu^{-2}\delta\theta_k)^{-1} \tag{2.5.7}$$

with $\theta_0 = |z_0|^2 - |z^*|^2 = \Gamma(z_0)$. Thus comparison of (2.5.6) and (2.5.7) yields $\Gamma(z_k) \le \theta_k$, $k \ge 0$. Finally, (2.5.2) and (2.5.3) follow from (2.4.8) and (2.4.9).

The complexity of the difference equation (2.3.4) makes it difficult to obtain more specific analytic results than those obtained in Theorem 2.5.1. Thus the remainder of this section is limited to the presentation and discussion of three example problems.

Example 1. Take $\delta = 1$ and let K be the convex hull of three points in 2-space, $(1, \nu), (-1, \nu), (0, 1 + \nu)$, where $\nu > 0$. Clearly $z^* = (0, \nu)$ and $|z^*| = \nu$. Simple inspection shows that the iterative process is finite ($z_1 = z^*$) if and only if z_0 is on the line segment connecting $(1, \nu)$ and $(-1, \nu)$. Moreover when the process is not finite z_k, $k \ge 1$, is determined by the scalar $\psi_k = |z_k^1|(z_k^2)^{-1}$. Thus the second order non-linear difference equation (2.3.4) may be replaced by a first order difference equation in ψ_k. It is not difficult to show that

$$\psi_{k+1} = \psi_k(1 - \nu\psi_k)(1 + \nu\psi_k + 2\psi_k^2)^{-1}, \quad k \ge 1. \tag{2.5.8}$$

For $\psi_k << 1$ this equation is approximated by $\tilde{\psi}_{k+1} = \tilde{\psi}_k(1 + 2\nu\tilde{\psi}_k)^{-1}$, an equation of the same form as (2.5.7). These observations and some tedious, but straightforward, computations lead to (the notation $o(\omega)$ means $\lim_{\omega \to 0} \omega^{-1} o(\omega) = 0$)

$$|z_k| - |z^*| = (2\nu k)^{-1} + o(k^{-1}), \tag{2.5.9}$$

$$|z_k - z^*| = (2\nu k)^{-1}\sqrt{1 + \nu^2} + o(k^{-1}) \tag{2.5.10}$$

Equation (2.5.9) demonstrates that is is impossible to obtain an upper bound on $|z_k| - |z^*|$ which approaches zero more rapidly than (const.) k^{-1}. For large k the upper bound in inequality (2.5.4) is conservative by a factor of sixteen. This factor can be traced to two sources each of which contributes a factor of four: in equation (2.4.15)

μ is an unsatisfactory estimate of $|z_k - s(-z_k)|$, in the derivation of inequality (2.4.6) the term $y \cdot (z^* - s(y))$ has been omitted. For this example the upper bound in inequality (2.5.5) is a poor estimate because it is order $k^{-\frac{1}{2}}$ rather than order k^{-1}.

It is also possible to show that

$$|z^*| - \gamma(z_k)|z_k| = (2\nu k)^{-1} + o(k^{-1}) \tag{2.5.11}$$

$$\sqrt{1 - \gamma(z_k)}\ |z_k| = k^{-\frac{1}{2}} + o(k^{-\frac{1}{2}}) \tag{2.5.12}$$

By comparing (2.5.11) with (2.5.9) and (2.5.12) with (2.5.10) it is seen that in Theorem 2.4.1, part iv) provides a reasonably good stopping criterion while part v) does not.

Example 2. Take $\delta = 1$ and let K be the convex hull of three points in 3-space, $(1, 0, \nu)$, $(-1, 0, \nu)$, $(0, 1, \nu)$, where $\nu > 0$. Thus $z^* = (0, 0, \nu)$ and $|z^*| = \nu$. The iterative process is much the same as in Example 1, the points $z_k \in K$, $k \geq 1$ being determined by the scalar $\psi_k = |z_k^1|(z_k^2)^{-1}$. The first order difference equation for ψ_k is (2.5.8) with $\nu = 0$. By using the fact that $\tilde{\psi}_k = \tilde{\psi}_0 (1 + 4\tilde{\psi}_0^2 k)^{-\frac{1}{2}}$ is the solution of $\tilde{\psi}_{k+1} = \tilde{\psi}_k(1 + 4\tilde{\psi}_k^2)^{-\frac{1}{2}}$ and that $(1 + 4\tilde{\psi}_k^2)^{\frac{1}{2}} \cong 1 + 2\psi_k^2$ for $\psi_k << 1$, the following results can be derived:

$$|z_k| - |z^*| = (8\nu k)^{-1} + o(k^{-1}) , \tag{2.5.13}$$

$$|z_k - z^*| = (2k^{\frac{1}{2}})^{-1} + o(k^{-\frac{1}{2}}), \tag{2.5.14}$$

$$|z^*| - \gamma(z_k)|z_k| = 3(8\nu k)^{-1} + o(k^{-1}), \tag{2.5.15}$$

$$\sqrt{1 - \gamma(z_k)}\,|z_k| = (2k)^{-\frac{1}{2}} + o(k^{-\frac{1}{2}}). \tag{2.5.16}$$

Equation (2.5.14) shows that the asymptotic behavior of $|z_k - z^*|$ matches the bound given in inequality (2.5.5), except for a multiplicative factor of eight. The bound given in inequality (2.5.4) is conservative by a multiplicative factor of 64. Comparison of (2.5.15) with

(2.5.13) and (2.5.16) with (2.5.14) shows that parts iv) and v) of Theorem 2.4.1 both provide reasonable stopping criteria.

Example 3 (The Hyperparaboloid Problem) Take $\delta = 1$ and in n-space let

$$K = \{z | z^1 \geq \nu + \tfrac{1}{2} \sum_{i=2}^{n} (z^i)^2 \lambda_i^{-1},\ z^1 \leq 2\nu\},\quad \nu, \lambda_2, \lambda_3, \ldots, \lambda_n > 0. \tag{2.5.17}$$

In the neighborhood of the optimum $z^* = (\nu, 0, \ldots, 0)$, ∂K is the elliptic hyperparaboloid $z^1 = \nu + \frac{1}{2}\sum_{i=2}^{n}(z^i)^2\lambda_i^{-1}$, where $\lambda_2, \ldots, \lambda_n$ are the principal radii of curvature at the vertex z^*. Many other convex sets K have a boundary surface which in the neighborhood of z^* can be closely approximated by such an elliptic hyperparaboloid (see Chapter I of Busemann [B6]). Thus this example is of general interest.

For $y^1 < 0$ and $\frac{1}{2}\sum_{i=2}^{n}(y^1)^{-2}\lambda_i(y^i)^2 < \nu$ it is easy to show that

$$s^1(y) = \nu + \tfrac{1}{2}\sum_{i=2}^{n}(y^1)^{-2}\lambda_i(y^i)^2,\quad s^i(y) = -(y^1)^{-1}\lambda_i y^i,\ i = 2, \ldots, n. \tag{2.5.18}$$

Let $\bar{\lambda} = \max_{i=2,\ldots,n}\{\lambda_i\}$ and assume the conditions

$$\zeta = \nu\sqrt{1 + 2\nu\bar{\lambda}^{-1}} > |y|,\quad -\nu \geq y^1 \tag{2.5.19}$$

are satisfied, which in turn imply $\frac{1}{2}\sum_{i=2}^{n}(y^1)^{-2}\lambda_i(y^i)^2 < \nu$. Thus (2.5.19) defines a set on which (2.5.18) is valid. Using this fact, $z^1 \geq \nu$ for $z \in K$, and (2.3.1) gives

$$\beta(z) = \frac{(z^1-\nu)^2 + \nu(z^1-\nu) + \sum_{i=2}^{n}(1+\frac{1}{2}(z^1)^{-1}\lambda_i)(z^i)^2}{(z^1-\nu)^2 + \sum_{i=2}^{n}(1+(z^1)^{-1}\lambda_i + (z^1)^{-2}\nu\lambda_i + (z^1)^{-2}\lambda_i^2)(z^i)^2 + (\frac{1}{2}\sum_{i=2}^{n}(z^1)^{-2}\lambda_i(z^i)^2)^2}, \tag{2.5.20}$$

$$z \neq z^*,\quad z \in K,\quad |z| < \zeta.$$

Because $z \in K$, $|z| < \zeta$ imply $z^1 \geq \nu$ and $\frac{1}{2}\sum_{i=2}^{n}(z^1)^2\lambda_i(z^i)^2 < \nu$ it follows that

$$\beta(z) \geq \frac{(z^1-\nu)^2 + \sum_{i=2}^{n}(z^i)^2}{(z^1-\nu)^2 + (1+\frac{5}{2}\bar{\lambda}\nu^{-1}+\bar{\lambda}^2\nu^{-2})\sum_{i=2}^{n}(z^i)^2} \geq \frac{1}{1+\frac{5}{2}\bar{\lambda}\nu^{-1}+\bar{\lambda}^2\nu^{-2}} = \underline{\beta}, \tag{2.5.21}$$

$$z \neq z^*, \; z \in K, \; |z| < \zeta .$$

Because $\beta(z^*) = 0$ this inequality implies that $\beta(z)$ is discontinuous on K at z^*.

By starting with (2.4.13) and repeating the derivation of Section 2.4 with $[z \cdot (z - s(-z))]|z - s(-z)|^2 = \beta(z) \geq \underline{\beta}$ it can be shown that

$$\Gamma(z_{k+1}) \leq \Gamma(z_k)(1 - \tfrac{1}{2}\underline{\beta}\delta), \qquad z_k \neq z^*. \tag{2.5.22}$$

For $z_k = z^*$, $\Gamma(z_{k+1}) = 0$ and (2.5.22) is trivially true. Thus $\Gamma(z_k) \leq \Gamma(z_0)(1 - \frac{1}{2}\underline{\beta})^k$, $k \geq 0$, $z_0 \in K$, $|z_0| < \zeta$. Using (2.4.8) and (2.4.9) this leads to

$$|z_k| - |z^*| \leq \tfrac{1}{2}\nu^{-1}\theta_0(1 - \tfrac{1}{2}\bar{\beta}\delta)^k, \tag{2.5.23}$$

$$|z_k - z^*| \leq \sqrt{\theta_0}\,(1 - \tfrac{1}{2}\bar{\beta}\delta)^{\frac{k}{2}} ,$$

where θ_0 is given as before in (2.5.1). Since $\underline{\beta} > 0$ inequalities (2.5.23) and (2.5.24) guarantee that the convergence of $\{|z_k|\}$ and $\{|z_k - z^*|\}$ is geometric. However, the guaranteed rate of convergence is not rapid if $\underline{\beta} << 1$, i.e., $\nu^2 << \bar{\lambda}^2$.

2.6 Numerical Results for BIP

Actual computations with BIP were carried out for Example 3 of Section 2.5 at the University of Michigan Computing Center. The programs were written in the MAD language and an IBM 7090 digital computer was used. In all the experiments $\nu = 1$, i.e., $z^* = (1, 0, \ldots, 0)$.

Also, the extent of K was increased beyond $z^1 = 2\nu$ so that (2. 5. 18) is valid even though (2. 5. 19) is violated. Data was obtained for n = 2 and 3 and various combinations of λ_2, λ_3, z_0^1, z_0^2, and z_0^3. The figures and tables included in this section contain an important part of these results.

In optimal control applications (see Chapter 5) the evaluation of a contact function is the most time-consuming part of the iterative procedure. Since such an evaluation is required on each iteration of BIP, the number of iterations to satisfy certain error criteria is used as a measure of the speed of convergence.

The results for n = 2 are presented first. Figure 2. 6. 1 shows the strong dependence of BIP on the parameter $\lambda_2 \nu^{-1}$, convergence becoming very slow as $\lambda_2 \nu^{-1} \to \infty$. Additional data for various performance measures and λ_2 values, given in Tables 2. 6. 1 through 2. 6. 4, illustrate further this dependence. It is interesting to note that $\beta(z_k)$ shown in Figure 2. 6. 2 is large immediately preceding major improvements in $|z_k| - |z^*|$. Table 2. 6. 5, which displays the case $\lambda_2 = 100$ for different points z_0, shows a somewhat random dependence of BIP on the initial point z_0. As a rough average, about 16 iterations are required for a decade of improvement in $|z_k| - |z^*|$ ($n = 2$, $\lambda_2 = 100$).

The results for n = 3 are presented in Figures 2. 6. 3, 2. 6. 4 and Tables 2. 6. 6 through 2. 6. 9. The behavior of $|z_k| - |z^*|$, $|z_k - z^*|$, $|z^*| - |z_k|\gamma(z_k)$, and $|s(-z_k) - z^*|$ shown in Figures 2. 6. 3 and 2. 6. 4 is typical of that observed in all computations with BIP. It can be stated that $|z^*| - |z_k|\gamma(z_k)$ decreases most rapidly, followed in order by $|z_k| - |z^*|$, $|z_k - z^*|$, and $|s(-z_k) - z^*|$. The results indicate that convergence is quite slow when $\nu << \bar{\lambda}$. Furthermore, compared with the parameter $\bar{\lambda}\nu^{-1}$, the ratio λ_2/λ_3 has little effect. By way of contrast, for iterative procedures of the gradient type [e. g. , Bl, El, Fl, F2, Nl, N3, N4], λ_2/λ_3 has an important influence.

Attempts to improve convergence for this example by taking $\delta \neq 1$ and choosing $\alpha_k \in I(z_k)$ in a variety of ways were unsuccessful.

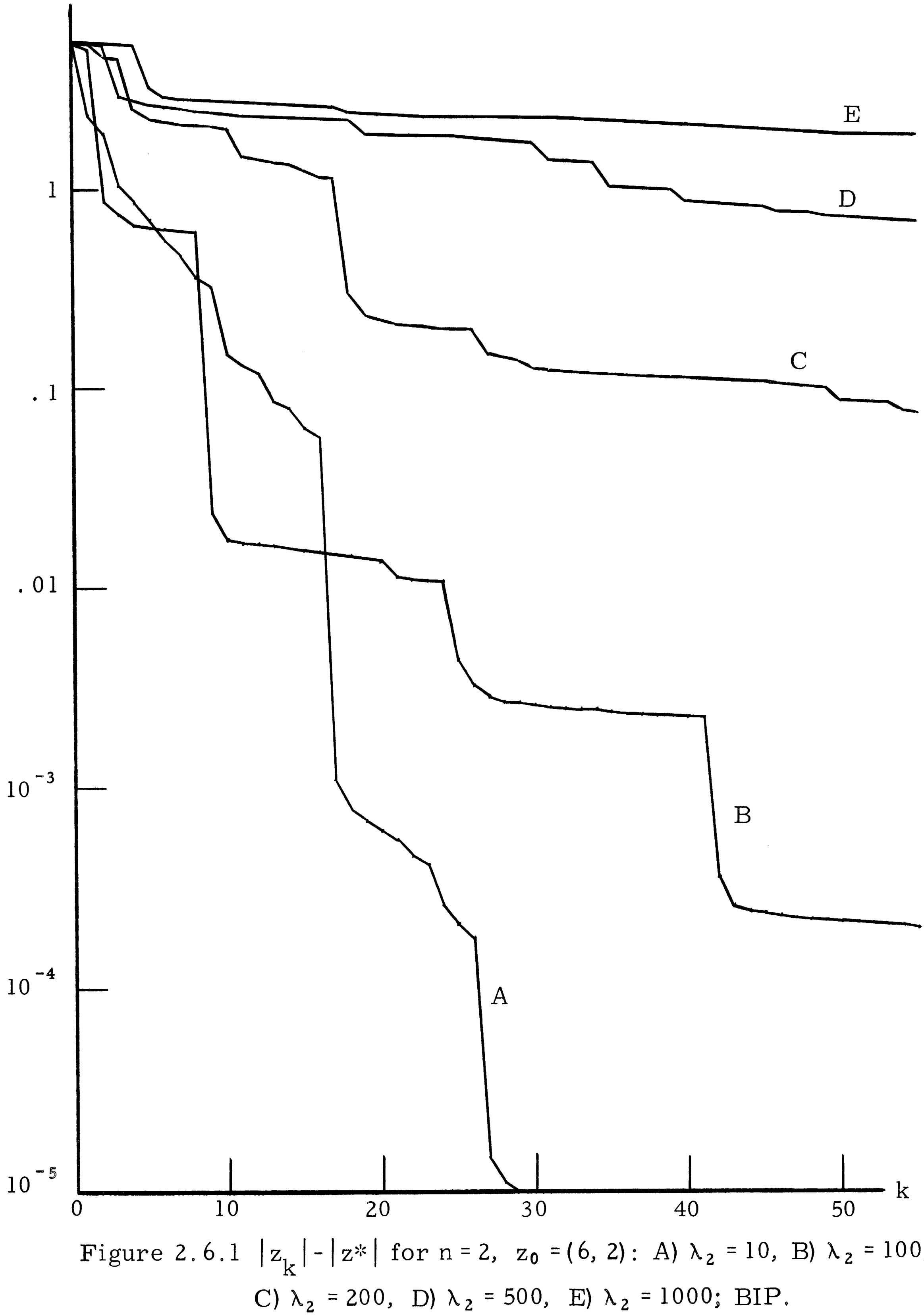


Figure 2.6.1 $|z_k| - |z^*|$ for $n = 2$, $z_0 = (6, 2)$: A) $\lambda_2 = 10$, B) $\lambda_2 = 100$, C) $\lambda_2 = 200$, D) $\lambda_2 = 500$, E) $\lambda_2 = 1000$; BIP.

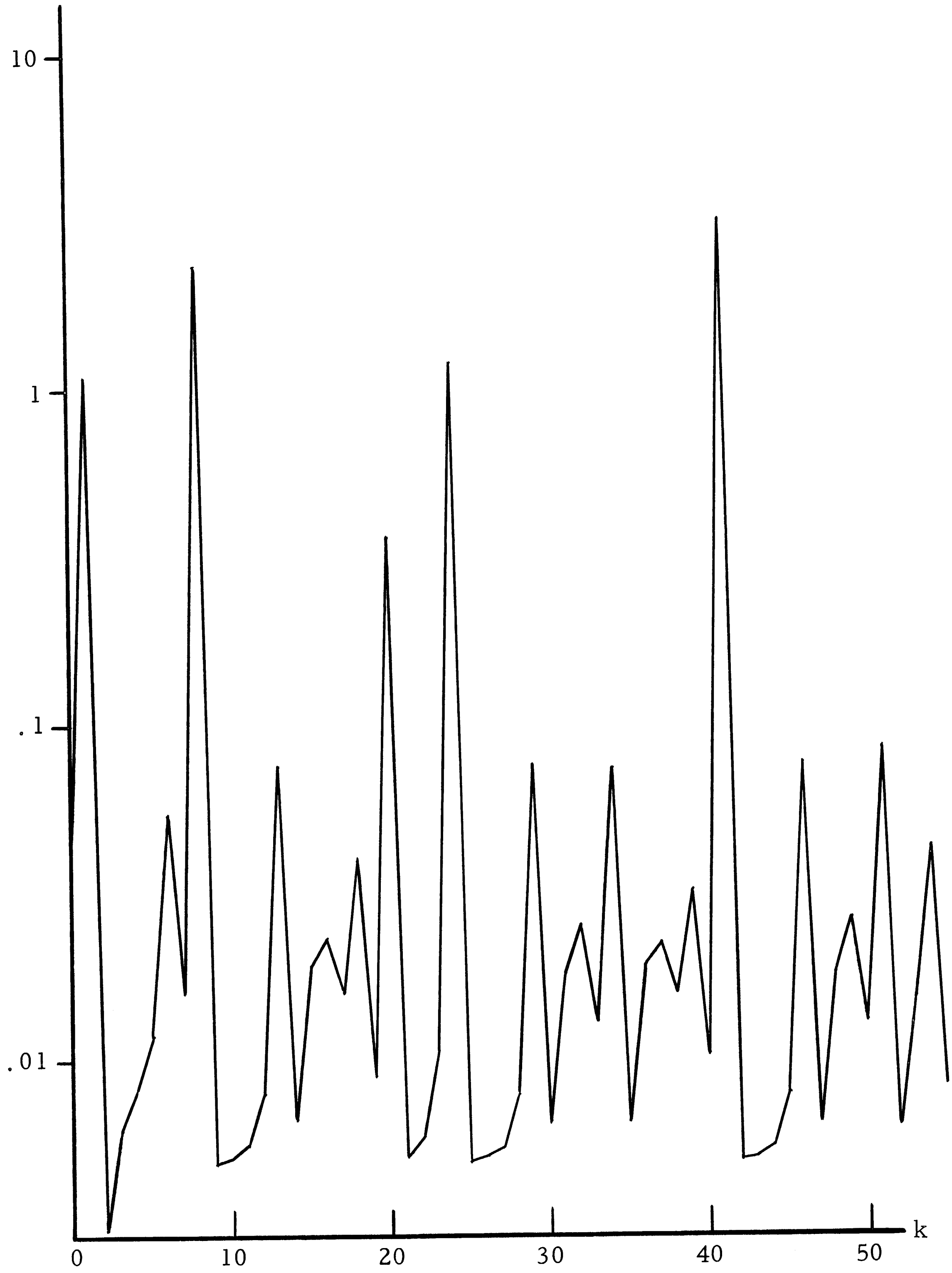


Figure 2.6.2 $\beta(z_k)$ for $n = 2$, $z_0 = (6, 2)$, $\lambda_2 = 100$; BIP.

Table 2.6.1 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$;

$n = 2$, $z_0 = (6, 2)$, BIP.

λ_2 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10	4	13	17	18	27	29	40
100	2	9	25	42	60	60	75
200	18	49	98	147	148	167	216
500	36	113	126	134	150	158	241
1000	69	96	200	215	281	376	443

Table 2.6.2 Number of iterations to satisfy $|z_k - z^*| \leq \epsilon$;

$n = 2$, $z_0 = (6, 2)$, BIP.

λ_2 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10	4	14	19	26	39	>40	
100	4	11	29	46	64	83	>98
200	18	49	99	151	166	213	>269
500	38	115	129	139	157	165	>300
1000	71	98	203	224	301	443	>500

Table 2.6.3 Number of iterations to satisfy $|z^*| - |z_k| \gamma(z_k) \leq \epsilon$, $n = 2$, $z_0 = (6, 2)$, BIP

λ_2 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10	0	9	16	16	23	26	33
100	0	1	8	8	24	41	59
200	0	3	17	49	99	146	146
500	0	2	34	112	112	125	141
1000	0	4	68	85	95	199	321

Table 2.6.4 Number of iterations to satisfy $|s(-z_k) - z^*| \leq \epsilon$; $n = 2$, $z_0 = (6, 2)$, BIP.

λ_2 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10	9	16	26	39	>40		
100	8	24	59	84	>98		
200	17	146	166	253	>269		
500	112	125	157	299	>300		
1000	85	199	392	487	>500		

Table 2. 6. 5 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$;
$n = 2$, $\lambda_2 = 100$, BIP.

z_0^1	z_0^2 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
6	1	18	19	44	44	62	102	137
6	1.5	5	36	57	79	90	100	129
6	2	2	9	25	42	60	60	75
6	2.5	13	24	34	34	44	49	49
6	3	4	11	30	30	51	69	92
6	3.5	11	34	34	34	34	49	49
6	4	18	22	34	34	59	81	81
6	4.5	10	10	22	37	38	71	89
6	5	3	15	47	47	47	47	47
2	6	5	14	33	61	62	115	136
3	5.57	25	67	72	89	122	160	160
4	4.9	6	16	28	59	93	106	134
4.5	4.44	10	20	62	73	88	113	137
5	3.87	3	23	50	50	71	92	121
5.5	3.12	8	40	68	102	120	136	175
6.2	1.25	10	21	23	36	36	36	36

Note: For $z_0 = (6, 2)$ and the last seven cases in the table, $|z_0|^2 \cong 40$.

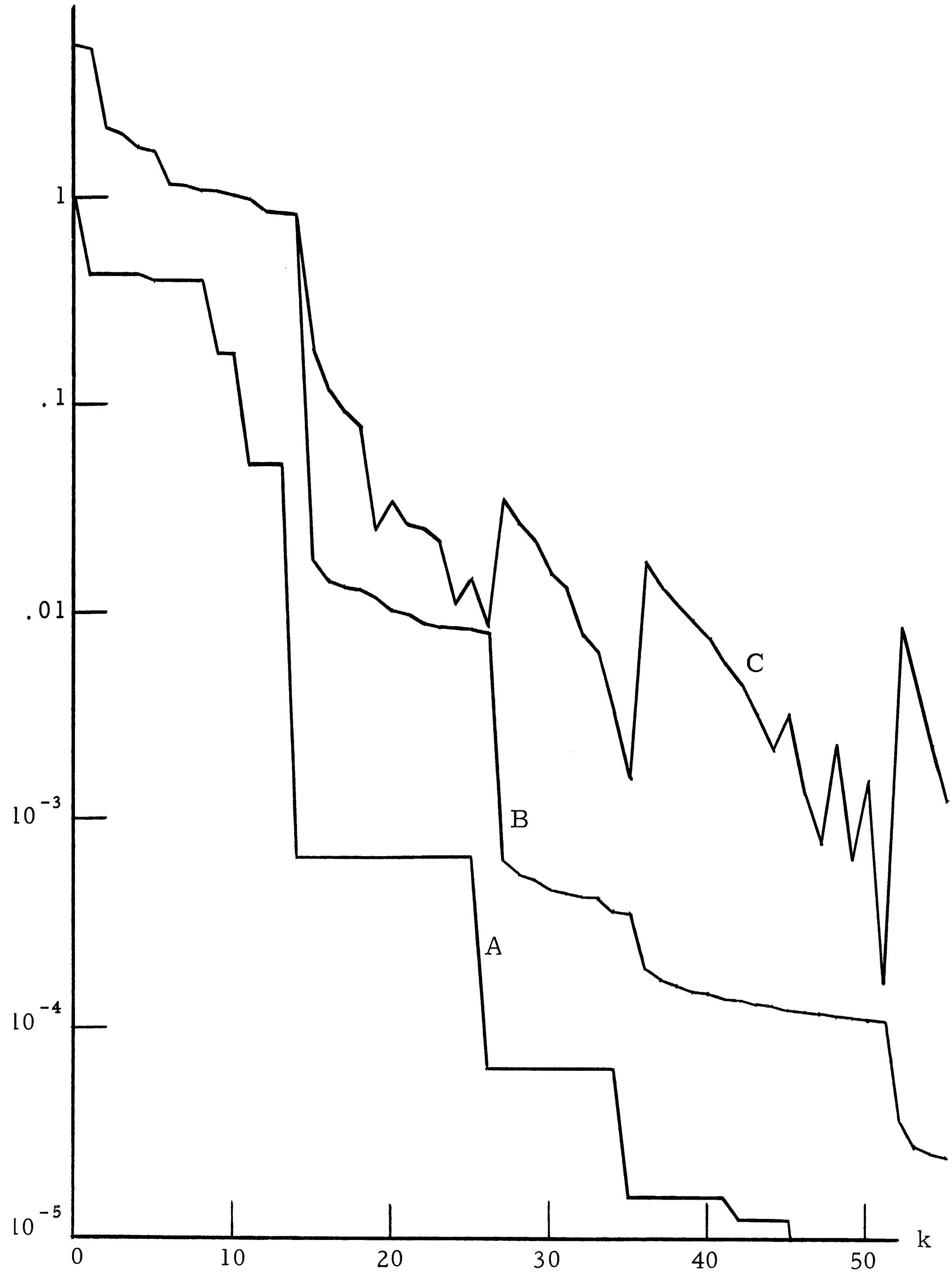


Figure 2.6.3 Results for $n = 3$, $z_0 = (6, 2, 2)$, $\lambda_2 = 100$, $\lambda_3 = 10$: A) $|z^*| - \max_{i \le k} |z_i| \gamma(z_i)$, B) $|z_k| - |z^*|$, C) $|z_k - z^*|$; BIP. For $k \le 14$, $|z_k - z^*| \cong |z_k| - |z^*|$.

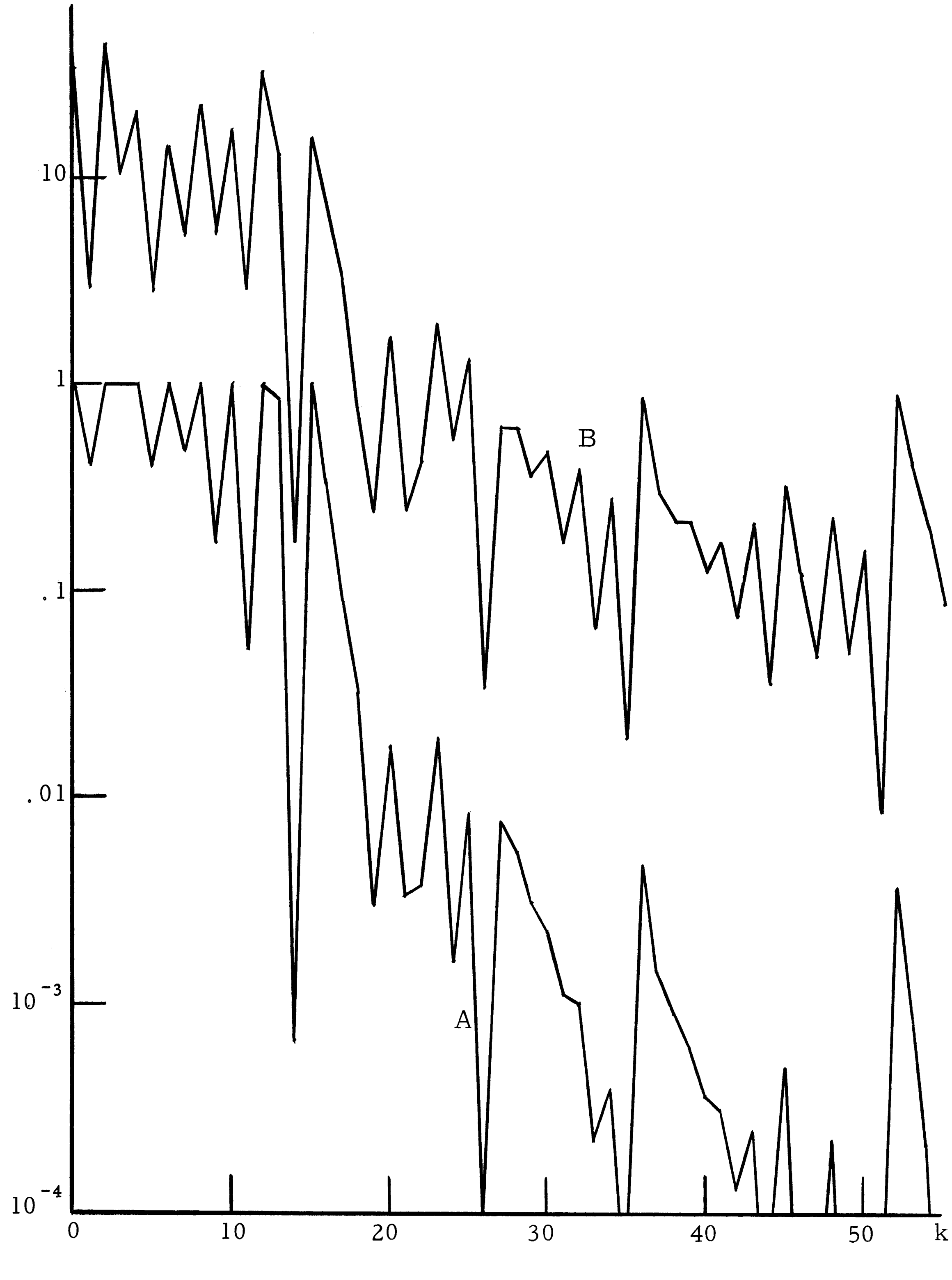


Figure 2.6.4 Results for n = 3, z_0 = (6, 2, 2), λ_2 = 100, λ_3 = 10: A) $|z^*| - |z_k|\gamma(z_k)$, B) $|s(-z_k) - z^*|$; BIP.

Table 2.6.6 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$;
$n = 3$, $z_0 = (6, 2, 2)$, BIP.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	1	2	3	3	5	5	5
10	10	4	9	21	28	31	38	41
100	100	10	19	31	59	59	74	111
1000	1000	82	82	126	216	250	290	340
100	1	2	6	24	35	46	46	57
100	10	11	15	21	27	52	73	81
100	50	20	33	78	124	124	151	151
100	90	15	38	38	58	112	137	137
1000	1	14	198	218	218	274	321	321
1000	10	83	174	207	229	267	298	359
1000	100	81	88	108	197	221	273	358

Table 2.6.7 Number of iterations to satisfy $|z_k - z^*| \leq \epsilon$;
$n = 3$, $z_0 = (6, 2, 2)$, BIP.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	1	3	4	4	5	5	> 5
10	10	4	11	25	30	40	> 41	
100	100	10	20	35	63	73	>111	
1000	1000	82	84	130	247	309	>340	
100	1	2	15	36	45	> 57		
100	10	11	17	26	47	64	92	>101
100	50	20	35	81	127	146	>151	
100	90	17	38	42	67	132	>137	
1000	1	16	199	232	238	318	>321	
1000	10	84	177	216	249	291	>359	
1000	100	81	90	112	203	259	337	>358

Table 2.6.8 Number of iterations to satisfy $|z^*| - |z_k| \gamma(z_k) \leq \epsilon$;
$n = 3$, $z_0 = (6, 2, 2)$, BIP.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	0	2	2	2	4	4	4
10	10	0	5	8	20	27	30	37
100	100	0	9	9	18	35	58	58
1000	1000	0	26	47	81	81	125	165
100	1	0	1	19	31	38	45	45
100	10	0	11	14	14	26	51	51
100	50	0	13	19	67	77	123	123
100	90	0	14	37	37	37	57	127
1000	1	0	13	41	146	197	241	255
1000	10	0	34	82	88	173	223	264
1000	100	0	9	80	80	87	107	196

Table 2.6.9 Number of iterations to satisfy $|s(-z_k) - z^*| \leq \epsilon$;
$n = 3$, $z_0 = (6, 2, 2)$, BIP.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	0	2	4	4	4	4	>5
10	10	5	20	30	40	>41		
100	100	9	35	73	>111			
1000	1000	81	125	165	292	>340		
100	1	1	23	45	56	>57		
100	10	14	26	51	92	>101		
100	50	19	77	123	150	>151		
100	90	37	37	127	136	>137		
1000	1	96	243	286	>321			
1000	10	159	241	295	>359			
1000	100	80	107	232	>358			

CHAPTER 3

THE IMPROVED ITERATIVE PROCEDURE IIP

The examples of Section 2. 5 and results of Section 2. 6 indicate that for many problems BP the convergence of BIP is not rapid. In Example 3, Section 2. 5 slow convergence is obtained with BIP for cases in which the surface ∂K at z^* has at least one principal radius of curvature large compared with $|z^*|$. For problems in which $S(-z^*)$ contains more than one point such as Examples 1 and 2, Section 2. 5, the convergence of BIP is especially poor. Since a large number of convex sets K which occur in practical problems either have a boundary surface near z^* approximately like that of Example 3 or have contact sets containing more than one point (K not strictly convex), it is important to seek ways to improve BIP.

By using more than one contact point at each iteration to gain information about ∂K, it is possible to develop an iterative procedure for solving BP which exhibits much more rapid convergence than BIP. Such an improved iterative procedure IIP is discussed in this chapter. First some background material on convex polyhedra is given after which IIP is stated and shown to converge. In the last two sections a sufficient condition for finite convergence of IIP and the results of many numerical computations are presented.

3.1 Convex Sets and Polyhedra

It is convenient to introduce here some basic definitions and results for convex sets and polyhedra which will be needed in the presentation of IIP. The proofs of all the results except Theorem 3. 1. 2 are omitted since they can be found in standard references such as Eggleston [E2] and Valentine [V2].

If X is an arbitrary set of points in E^n, then the convex hull of X, written ΔX, is the set of points which is the intersection of all the convex sets that contain X. ΔX is convex, and a necessary and sufficient condition that X be convex is that $X = \Delta X$. Furthermore, if X is compact, then ΔX is compact.

The convex hull of a finite number of points from E^n is called a convex polyhedron. This set is compact and may be viewed as the intersection of a finite number of closed half-spaces. Thus it has a representation in terms of a finite set of linear equations and/or inequalities. Given $y_1, y_2, \ldots, y_m \in E^n$, the convex polyhedron $\Delta\{y_1, y_2, \ldots, y_m\}$ is identical with the set of points of the form $y = \sum_{i=1}^{m} \sigma_i y_i$, where $\sum_{i=1}^{m} \sigma_i = 1$ and $\sigma_i \geq 0$ $(i = 1, 2, \ldots, m)$. Each such point y is said to be a convex combination of $y_1, y_2, \ldots, y_m$.

A point $y \in E^n$ is said to be an extreme point of a convex set X if $y \in X$ and there do not exist two distinct points $y_1, y_2 \in X$ such that $y \in \Delta\{y_1, y_2\}$, $y \neq y_1$, $y \neq y_2$. The set of extreme points of a convex polyhedron is finite and is called the skeleton of the convex polyhedron. Furthermore, a convex polyhedron is the convex hull of its skeleton.

The dimension of a convex set $X \subset E^n$, written dim X, is the largest integer m such that X contains m + 1 points $y_1, y_2, \ldots, y_{m+1}$ for which the collection of vectors $\{y_1 - y_{m+1}, \ldots, y_m - y_{m+1}\}$ is linearly independent. It follows that dim $X \leq n$. A non-empty set in E^n is called a linear manifold if it consists of a single point or if for every y_1, y_2 in the set, $y_1 \neq y_2$, the line $L(y_1; y_2)$ is in the set. If dim X = m, then X is a subset of an m-dimensional linear manifold (i. e., a point if m = 0, a line if m = 1, a plane if m = 2, a hyperplane if m = n - 1, the whole space E^n if m = n). The relative interior and relative boundary of a convex set X having dimension m are defined to be the interior and boundary of X relative to the m-dimensional linear manifold which contains X.

A convex polyhedron in E^n having dimension m is called an m-polyhedron. The relative boundary of an m-polyhedron is the union of a finite number of (m - 1)-polyhedra. Furthermore, the skeleton of each (m - 1)-polyhedron is a subset of the skeleton of the m-polyhedron.

A simplex is a special case of a convex polyhedron and is defined as follows: an m-dimensional simplex, or more briefly an m-simplex, in E^n ($m \leq n$) is the convex hull of m + 1 points from E^n which do not lie on a linear manifold of dimension m - 1. The set of m + 1 points is the skeleton of the m-simplex.

THEOREM 3.1.1 (Carathéodory [C1]) Let $X \subset E^n$. If $y \in \Delta X$, there is a set of m points $y_1, y_2, \ldots, y_m$ all belonging to X with $m \leq n + 1$ such that y is contained in the (m - 1)-simplex $\Delta\{y_1, y_2, \ldots, y_m\}$.

The following theorem is a collection of some rather obvious results on convex polyhedra which will be useful in the sequel.

THEOREM 3.1.2 Given m points $y_1, y_2, \ldots, y_m \in E^n$, consider the convex polyhedron $H = \Delta\{y_1, y_2, \ldots, y_m\}$. Let the dimension of H be q. Then: i) $q \leq \min\{m - 1, n\}$; ii) the skeleton of H, which is a subset of $\{y_1, y_2, \ldots, y_m\}$, contains at least q + 1 points; iii) if q = n, ∂H is the union of a finite number of (n - 1)-polyhedra; iv) if q < n, $\partial H = H$ (a q-polyhedron); v) if $m \geq n$ and $\Delta_1, \Delta_2, \ldots, \Delta_{\overline{m}}$, $\overline{m} = \frac{m!}{n!(m-n)!}$, denote the convex polyhedra formed by the convex hull of n points chosen from $\{y_1, y_2, \ldots, y_m\}$, then $\partial H \subset \bigcup_{j=1}^{\overline{m}} \Delta_j$; vi) if $m \geq n$ and $y \in \partial H$, then y has the representation $y = \sum_{i=1}^{m} \sigma_i y_i$ where $\sum_{i=1}^{m} \sigma_i = 1$, $\sigma_i \geq 0$ (i = 1, 2, . . . , m), and at least m - n values of σ_i, $1 \leq i \leq m$, are 0; vii) if $m \geq n + 1$ and $\Delta_1, \Delta_2, \ldots, \Delta_{\overline{\ell}}$, $\overline{\ell} = \frac{m!}{(n+1)!(m-n-1)!}$, denote the convex polyhedra formed by the convex hull of n + 1 points chosen from $\{y_1, y_2, \ldots, y_m\}$, then $H = \bigcup_{j=1}^{\overline{\ell}} \Delta_j$.

Proof: Clearly $q \leq n$ so i) is true for $n \leq m - 1$. Consider $n > m - 1$. If $q > m - 1$, then there exist $x_1, x_2, \ldots, x_{m+1} \in H$ such that the set of vectors $\{x_1 - x_{m+1}, \ldots, x_m - x_{m+1}\}$ is linearly independent. Every $y \in H$ has the representation $y = \sum_{i=1}^{m} \sigma_i y_i$, $\sum_{i=1}^{m} \sigma_i = 1$. That is, $y = y_1(1 - \sum_{i=2}^{m} \sigma_i) + \sum_{i=2}^{m} \sigma_i y_i = y_1 + \sum_{i=2}^{m} \sigma_i (y_i - y_1)$. Thus each vector $x_j - x_{m+1}$ $(j = 1, 2, \ldots, m)$ can be written as a linear combination of $y_i - y_1$ $(i = 2, 3, \ldots, m)$, contradicting the fact that the $x_j - x_{m+1}$ are linearly independent. Hence, $q \leq m - 1$ which completes the proof of i).

Consider ii). Suppose there is a point $x \neq y_i$, $1 \leq i \leq m$, which is an extreme point of H. Since $x \in H$, $x = \sum_{i=1}^{m} \sigma_i y_i$ where $\sum_{i=1}^{m} \sigma_i = 1$, $\sigma_i \geq 0$ $(i = 1, 2, \ldots, m)$. But $x \neq y_i$, $1 \leq i \leq m$, implies $\sigma_i < 1$ $(i = 1, 2, \ldots, m)$ and at least two σ_i are > 0. For definiteness, suppose $\sigma_i > 0$. From $(1 - \sigma_1)^{-1} \sum_{i=2}^{m} \sigma_i = 1$ and $(1 - \sigma_1)^{-1} \sigma_i \geq 0$ $(i = 2, 3, \ldots, m)$, it follows that $\bar{y} = (1 - \sigma_1)^{-1} \sum_{i=2}^{m} \sigma_i y_i \in H$. Then $x = \sigma_1 y_1 + (1 - \sigma_1)\bar{y}$ where $x \neq y_1$ and $x \neq \bar{y}$, violating the assumption that x is an extreme point. Let p be the number of points in the skeleton of H. Suppose $p \leq q$. Since $q \leq n$, $p \leq n$. Then H is the convex hull of p points, and part i) implies $\dim H \leq p - 1$. But $p \leq q$ yields $q = \dim H \leq q - 1$, a contradiction. Thus ii) is established.

If $q = n$, ∂H equals the relative boundary of H and thus is the union of a finite number of $(n - 1)$-polyhedra. This proves iii). In iv) the fact that H is a subset of a q-dimensional linear manifold, $q < n$, implies that if $y \in H$ and $\epsilon > 0$ then $N(y;\epsilon)$ contains a point not in H. Thus $\partial H = H$.

Consider v). By iii) and iv) ∂H is the union of a finite number of convex polyhedra each having dimension $\leq n - 1$. From the remarks following the definition of m-polyhedron, the set of extreme points of each of these convex polyhedra is a subset of $\{y_1, y_2, \ldots, y_m\}$. Thus given $y \in \partial H$ there is a set $X \subset \{y_1, y_2, \ldots, y_m\}$ such that $y \in \Delta X$ and X is contained in a linear manifold of dimension $n - 1$. Since $\dim X \leq n - 1$, Theorem 3.1.1 applied to X yields: there exist $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p \in X$ with $p \leq n$ such that $y \in \Delta\{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p\}$. But for $p \leq n$ and any set

$\{\bar{y}_1 \bar{y}_2, \ldots, \bar{y}_p\}$ of points from $\{y_1, y_2, \ldots, y_m\}$, $\Delta\{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p\} \subset \bigcup_{j=1}^{\bar{m}} \Delta_j$ and therefore v) is true.

Part vi) follows at once from v) since $y \in \partial H$ implies y is contained in at least one Δ_j, $1 \le j \le \bar{m}$. Thus y may be written as a convex combination of n points from $\{y_1, y_2, \ldots, y_m\}$ and the remaining σ_i in $\sum_{i=1}^{m} \sigma_i y_i$ may be set equal to 0.

Consider vii). Let $X = \{y_1, y_2, \ldots, y_m\}$. Then Theorem 3.1.1 yields: given $y \in \Delta X = H$ there exist $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p \in X$ with $p \le n+1$ such that $y \in \Delta\{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p\}$. If Δ_j, $1 \le j \le \bar{\ell}$, are defined as in vii), then $\Delta\{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p\} \subset \bigcup_{j=1}^{\bar{\ell}} \Delta_j$ for any set $\{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p\}$, $p \le n+1$, of points from X. Hence $H \subset \bigcup_{j=1}^{\bar{\ell}} \Delta_j$. Now suppose $x \in \bigcup_{j=1}^{\bar{\ell}} \Delta_j$. Then $x \in \Delta_j$ for some j, $1 \le j \le \bar{\ell}$, which means x is a convex combination of n + 1 points from X. By giving the remaining m - n - 1 points in X a zero coefficient, x can be written as a convex combination of all the points in X. Hence $x \in H$ and $\bigcup_{j=1}^{\bar{\ell}} \Delta_j \subset H$. This completes the proof.

3.2 The Subproblem 1

On each iteration of IIP the following minimization problem, which is a special case of BP, will occur.

SUBPROBLEM 1 Given: H, the convex hull of m points $y_1, y_2, \ldots, y_m$ from E^n. Find: a point $y^* \in H$ such that $|y^*| = \min_{y \in H} |y|$.

Since H is compact and $|y|$ is a continuous function of y, a solution y^* exists. Let $P_H(x)$, $x \neq 0$, be the support hyperplane of H with outward normal x. Then:

THEOREM 3.2.1 (Solution Properties for Subproblem 1) i) y^* is unique; ii) $|y^*| = 0$ if and only if $0 \in H$; iii) for $|y^*| > 0$, $y^* \in \partial H$; iv) for $|y^*| > 0$, $y = y^*$ if and only if $y \in P_H(-y) \cap H$; v) if $m \ge n$ and $y^* \in \partial H$, y^* has a representation $y^* = \sum_{i=1}^{m} \sigma_i y_i$ where $\sum_{i=1}^{m} \sigma_i = 1$,

$\sigma_i \geq 0$ $(i = 1, 2, \ldots, m)$, and at least $m - n$ values of σ_i $1 \leq i \leq m$, are 0;
vi) if $m \geq n$, $|y^*| > 0$, and $y^* = \sum_{i=1}^{\bar{n}} \bar{\sigma}_i \bar{y}_i$ where $\bar{n} \leq m$, $\sum_{i=1}^{\bar{n}} \bar{\sigma}_i = 1$, $\bar{\sigma}_i > 0$ $(i = 1, 2, \ldots, \bar{n})$, and $\bar{y}_i$ $(i = 1, 2, \ldots, \bar{n})$ are vectors from $\{y_1, y_2, \ldots, y_m\}$, then $\bar{y}_i$ $(i = 1, 2, \ldots, \bar{n})$ are contained in $P_H(-y^*)$.

Proof: Parts i) through iv) follow by the same arguments used in proving Theorem 2.2.1 and v) comes directly from part vi) of Theorem 3.1.2. Consider vi). By iv) $y^* \in P_H(-y^*)$ so vi) is clearly true for $\bar{n} = 1$. Thus take $1 < \bar{n} \leq m$. Suppose any $\bar{y}_i$, say $\bar{y}_1$, $\notin P_H(-y^*)$. Then $(\bar{y}_1 - y^*) \cdot y^* \neq 0$. This, $\bar{y}_1 \in H$, and the fact that $P_H(-y^*)$ is a support hyperplane of H with outward normal $-y^*$ imply $(\bar{y}_1 - y^*) \cdot y^* > 0$. Since $y^* = \sum_{i=1}^{\bar{n}} \bar{\sigma}_i \bar{y}_i$ where $\sum_{i=1}^{\bar{n}} \bar{\sigma}_i = 1$, $\bar{n} \geq 2$, and $\bar{\sigma}_i > 0$ $(i = 1, 2, \ldots, \bar{n})$, $\bar{\sigma}_1 < 1$ and $y^* = \bar{\sigma}_1 \bar{y}_1 + (1 - \bar{\sigma}_1)\tilde{y}$ where $\tilde{y} = (1 - \bar{\sigma}_1)^{-1} \sum_{i=2}^{\bar{n}} \bar{\sigma}_i \bar{y}_i$. From $(1 - \bar{\sigma}_1)^{-1} \sum_{i=2}^{\bar{n}} \bar{\sigma}_i = 1$ and $(1 - \bar{\sigma}_1)^{-1} \bar{\sigma}_i > 0$ $(i = 2, 3, \ldots, \bar{n})$, it follows that $\tilde{y} \in H$ which implies $(\tilde{y} - y^*) \cdot y^* \geq 0$. But $y^* = \bar{\sigma}_1 \bar{y}_1 + (1 - \bar{\sigma}_1)\tilde{y}$, $0 < \bar{\sigma}_1 < 1$, only if there exists $\epsilon > 0$ such that $y^* - \tilde{y} = \epsilon(\bar{y}_1 - y^*)$. Then $(\bar{y}_1 - y^*) \cdot y^* > 0$ yields $\frac{1}{\epsilon}(y^* - \tilde{y}) \cdot y^* > 0$. This leads to $(\tilde{y} - y^*) \cdot y^* < 0$, which contradicts the earlier result and completes the proof.

It is important to note the distinction between BP and Subproblem 1, both of which are quadratic programming problems on a compact, convex constraint set. The set K in BP is described only by a contact function $s(\cdot)$ of K whereas the convex polyhedron H in Subproblem 1 is the convex hull of m known points. Thus Subproblem 1 is much simpler than BP. It is shown in Section 3.6 that Subproblem 1 is amenable to solution by standard quadratic programming techniques.

3.3 The Improved Iterative Procedure IIP

In this section the improved iterative procedure for solving BP is described.

First let $s(\cdot)$ be a specific contact function of the set K in BP and consider:

$$\mu(z) = |z|^{-1} z \cdot s(-z) \qquad , \; z \neq 0 \qquad (3.3.1)$$
$$= 0, \qquad\qquad z = 0 \; .$$

Thus the function $\mu(\cdot)$, which is closely related to $\gamma(\cdot)$ of (2.3.2), is defined on K. Geometrically, $|\mu(z)|$, $z \neq 0$, is the Euclidean distance from the origin to the support hyperplane $P(-z)$ of K. The following properties also hold.

THEOREM 3.3.1 <u>Let</u> K <u>be the set described in</u> BP <u>and restrict</u> z <u>to</u> K. <u>Then</u>: i) $\mu(z) \leq |z|\gamma(z) \leq |z^*|$; ii) <u>if</u> $\mu(z) \geq 0$, $\mu(z) = |z|\gamma(z)$; iii) $\mu(z^*) = |z^*|$; iv) <u>if</u> $0 \notin K$ <u>and</u> $\mu(z) = |z^*|$, $S(-z) = S(-z^*)$; v) <u>if</u> $0 \notin K$, $\mu(z)$ <u>is continuous on</u> K.

Proof: In this paragraph z always denotes a point in K. The inequality $|z|\gamma(z) \leq |z^*|$ was shown in the proof of Theorem 2.4.1. Part ii) and the remaining result in i) follow from equations (2.3.2) and (3.3.1). Consider iii). If $0 \in K$, $z^* = 0$ and clearly $\mu(z^*) = |z^*| = 0$. If $0 \notin K$, part v) of Theorem 2.3.1 yields $\gamma(z^*) = 1$. But $\gamma(z) > 0$ implies $\mu(z) = |z|\gamma(z) > 0$ and thus $\mu(z^*) = |z^*|\gamma(z^*) = |z^*|$. In iv) $0 \notin K$ implies $|z| \geq |z^*| > 0$ so $P(-z)$, the support hyperplane of K with outward normal $-z$, is defined. But $P(-z) = \{x : x \cdot |z|^{-1}(-z) = s(-z) \cdot |z|^{-1}(-z) = -\mu(z)\}$ is also the support hyperplane of $\bar{N}(0;\mu(z))$ with outward normal z and contact point $\mu(z)|z|^{-1}z$. Thus if $\mu(z) = |z^*|$, $P(-z)$ is a (separating) support hyperplane for K and $\bar{N}(0;|z^*|)$. This implies $P(-z) = P(-z^*)$ and $S(-z) = S(-z^*)$. Consider v). From (3.3.1) and the continuity of the support function $\eta(y) = y \cdot s(y)$, $y \in E^n$, it follows that $\mu(z)$ is continuous except possibly at $z = 0$. For $0 \notin K$, $|z| > 0$ for all $z \in K$ so the proof is complete.

Now consider:

<u>The Improved Iterative Procedure IIP</u> Let $s(\cdot)$ be an arbitrary

contact function of the set K specified in BP. Take $z_0 \in K$ and choose a positive integer p. Then a sequence of vectors $\{z_k\}$, $k = 0, 1, 2, \ldots,$ in E^n is generated as follows:

<u>Step 1</u> Select any p vectors $y_1(k), y_2(k), \ldots, y_p(k)$ in K and let

$$Y_k = \{y_1(k), y_2(k), \ldots, y_p(k)\}, \tag{3.3.2}$$

$$H_k = \Delta\{y_1(k), y_2(k), \ldots, y_p(k), \; s(-z_k), \; z_k\} \; . \tag{3.3.3}$$

<u>Step 2</u> Solve Subproblem 1 on the convex polyhedron H_k: find $z_{k+1} \in H_k$ such that $|z_{k+1}| = \min_{z \in H_k} |z|$.

Steps 1 and 2 constitute one iteration, called iteration k, of IIP. Note that IIP differs from BIP (with $\delta = 1$) only in the fact that z_{k+1} is obtained by minimizing over H_k instead of $\Delta\{s(-z_k), z_k\}$.

There are a great variety of selection rules for choosing the elements of Y_k in Step 1 and each choice of a selection rule gives a different version of IIP. As is discussed in Section 3.5 the function $\mu(\cdot)$ forms the basis of selection for several versions of IIP which exhibit good convergence.

3.4 Convergence Theorem for IIP

THEOREM 3.4.1 <u>Consider the sequence</u> $\{z_k\}$ <u>generated by</u> IIP. <u>For</u> $k \geq 0$ <u>and</u> $k \to \infty$: i) $z_k \in K$; ii) <u>the sequence</u> $\{|z_k|\}$ <u>is decreasing</u> $(|z_k| \geq |z_{k+1}|)$, $|z_k| \to |z^*|$, <u>and</u> $|z_k| = |z_{k+1}|$ <u>implies</u> $z_k = z^*$; iii) $z_k \to z^*$; iv) $|z_k|\gamma(z_k) \leq |z^*|$ <u>and</u> $|z_k|\gamma(z_k) \to |z^*|$; v) $|z_k - z^*| \leq \sqrt{1 - \gamma(z_k)}\,|z_k|$ <u>and</u> $\sqrt{1 - \gamma(z_k)}\,|z_k| \to 0$; vi) $|s(-z_k) - z^*| \leq |s(-z_k) - \gamma(z_k) z_k|$. <u>Furthermore</u>: vii) $\mu(z_k) \leq |z_k|\gamma(z_k)$; vii) <u>if</u> $0 \notin K$, $\mu(z_k) \to |z^*|$.

It should be observed that parts i) through vi) are identical to the corresponding parts of Theorem 2.4.1 (convergence theorem for BIP).

The remarks in the two paragraphs following the statement of Theorem 2.4.1 also apply here.

Proof of Theorem 3.4.1: Since $Y_k \subset K$ and $s(-z_k) \in K$, it follows from the convexity of K and the definition of convex hull that $z_k \in K$ implies $H_k \subset K$ and $z_{k+1} \in K$. Thus by induction $z_0 \in K$ proves part i).

As in the proof of Theorem 2.4.1, the inequalities (2.4.1) through (2.4.7) and $z_k \in K$ yield the inequalities in iv), v), and vi). Part i) of Theorem 3.3.1 and i) prove vii).

Consider ii). Since IIP differs from BIP (with $\delta = 1$) only in the fact that z_{k+1} is obtained by minimizing over H_k instead of $\Delta\{s(-z_k), z_k\}$, it follows from $\Delta\{s(-z_k), z_k\} \subset H_k$ that $|z_{k+1}|$ using IIP $\le |z_{k+1}|$ using BIP ($\delta = 1$). By comparison of the sequences $\{|z_k|\}$, the first two results in ii) are a consequence of the corresponding results in part ii) of Theorem 2.4.1.

Now suppose $|z_k| = |z_{k+1}|$ and $z_k \neq z^*$. Since $|z_k| \ge |z^*| \ge 0$, $z_k \neq z^*$ implies $|z_k| > 0$ so that support hyperplanes $P(-z_k)$ and $P_{H_k}(-z_k)$ of K and H_k respectively are defined. Clearly $s(-z_k) \in P(-z_k) = Q(s(-z_k); z_k)$ and part iv) of Theorem 3.2.1 yields $z_{k+1} \in P_{H_k}(-z_{k+1}) = Q(z_{k+1}; z_{k+1})$. From $z_k, z_{k+1} \in H_k$, $|z_k| = |z_{k+1}| = \min_{z \in H_k} |z|$, and part i) of Theorem 3.2.1, it follows that $z_k = z_{k+1}$ and $Q(z_k; z_k) = Q(z_{k+1}; z_{k+1})$. Then $Q(z_k; z_k)$ is the support hyperplane of H_k with outward normal $-z_k$ and $s(-z_k) \in H_k \subset Q^-(z_k; z_k)$. Furthermore, $H_k \subset K$ implies $Q(z_k; z_k) \subset Q^-(s(-z_k); z_k)$. The last two statements can be true only if $Q(z_k; z_k) = Q(s(-z_k); z_k)$. Hence $z_k \in P(-z_k)$ which by i) and part iv) of Theorem 2.2.1 implies $z_k = z^*$. This contradiction completes the proof of ii).

Let the function $\Gamma(z)$ satisfying (2.4.8) and (2.4.9) be introduced again. Part ii) and (2.4.8) imply that $\{\Gamma(z_k)\}$ is decreasing and $\Gamma(z_k) \to 0$. By (2.4.9) this proves iii). The remaining results in parts iv) and v) follow from the known value of $\gamma(z^*)$, the continuity of

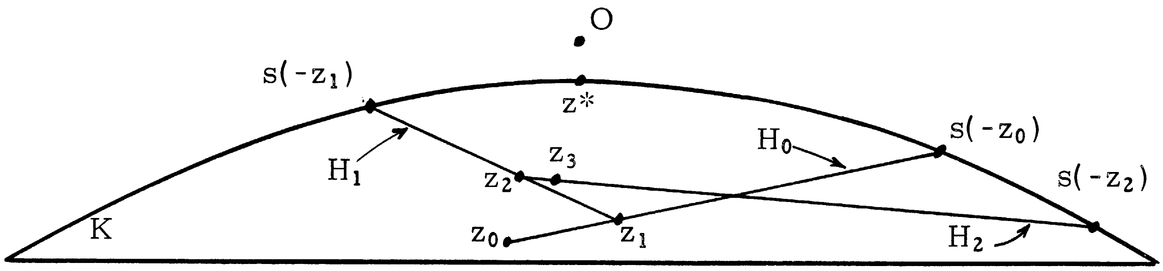
$\gamma(\cdot)$, and iii). Parts iii) and v) of Theorem 3.3.1 and iii) give viii). Thus the proof is complete.

3.5 Selection Rules for IIP

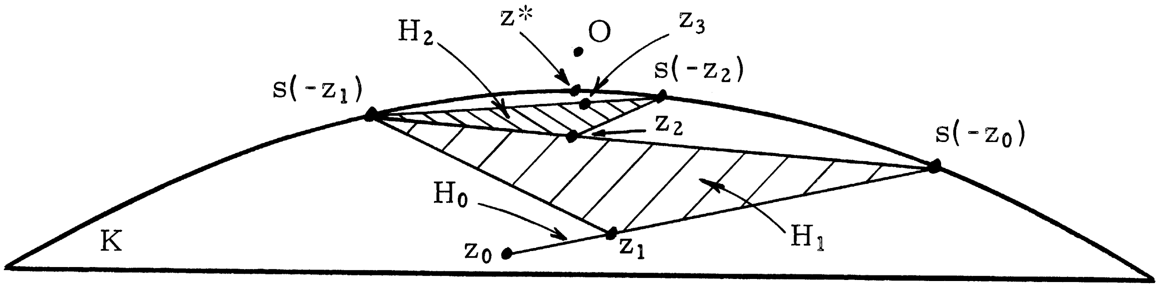
In this section several rules for selecting the vectors $y_1(k), y_2(k), \ldots, y_p(k) \in K$ in Step 1 of iteration k of IIP are presented. Each selection rule yields a particular version of IIP and it is desired that the iterative procedure converge much more rapidly than BIP.

Consider the case $z^* \in \partial K$ which is most important in applications. If after a few iterations of IIP the surface ∂H_k in the vicinity of z_{k+1} closely approximates ∂K in the vicinity of z^*, then it is likely that IIP will exhibit improved convergence. If ∂H_k is to approximate ∂K, the dimension of H_k must be sufficiently large, namely n, and Y_k must include boundary points of K. To illustrate these remarks, consider Figure 3.5.1 which is Example 3 of Section 2.5 with $n = 2$, $\nu = 1$, $\lambda_2 = 1$. In Figure 3.5.1(a) BIP is shown, where $\dim H_k = 1$ and convergence is slow. In Figure 3.5.1(b) IIP with Selection Rule A (to be described subsequently) is shown, where $\dim H_k \leq 2$ and convergence is notably improved. An even more startling improvement is exhibited in Figure 3.5.2 which is Example 1 of Section 2.5 with $\nu = 1$. Theorem 3.7.1 shows that when K is a convex polyhedron, IIP (with a suitable selection rule and contact function) converges in a finite number of iterations. Furthermore, the extensive numerical results of Section 3.8 provide strong evidence that IIP is far superior to BIP.

Let the p points in Y_k be contact points of K. Observe that ∂H_k is a better local approximation to ∂K for larger values of p. However, the larger p is, the more difficult it is to solve the Subproblem 1 in Step 2 of IIP. The computational results of Section 3.8 indicate that convergence is good for $p = n$ and little improvement is obtained for $p > n$. The desirability of choosing $p = n$ is also evident from the



(a) BIP.

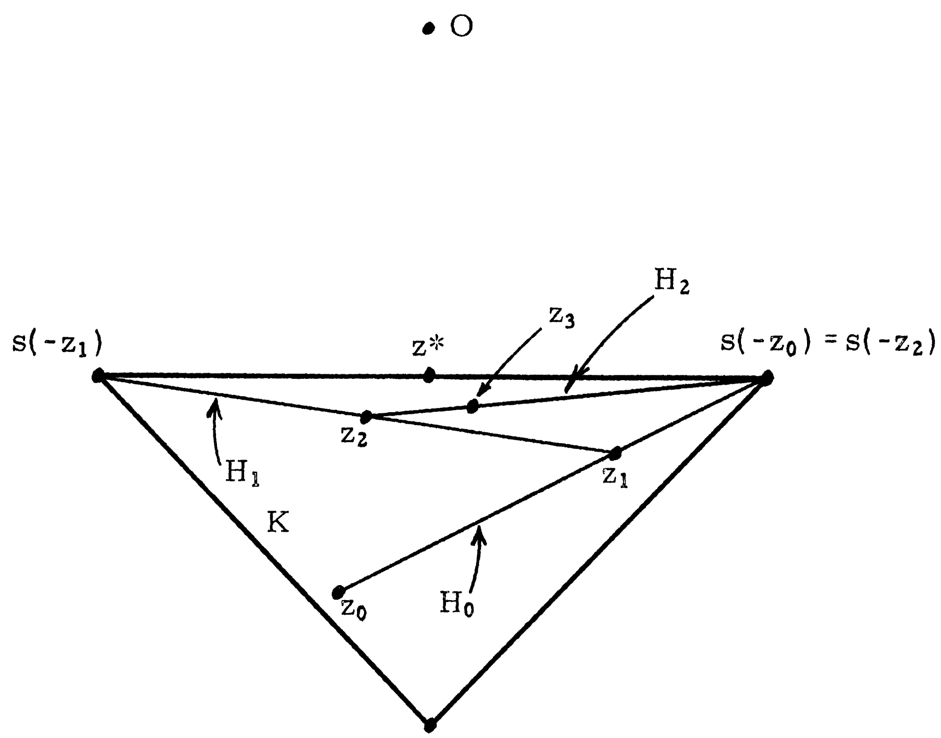


(b) IIP Selection Rule A, $p = 2$.

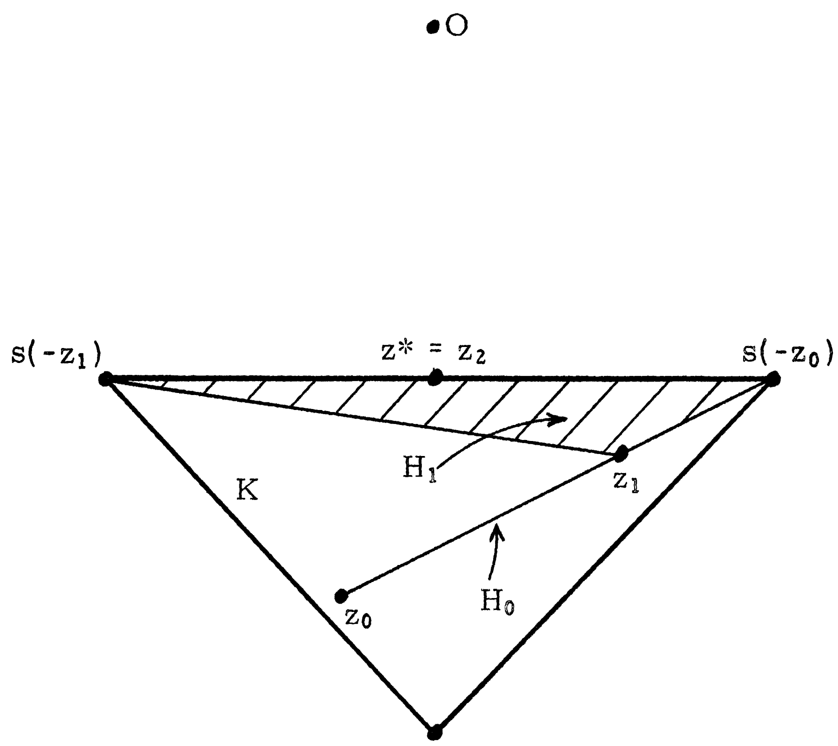
Figure 3.5.1 Example 3 of Section 2.5, $n = 2$, $\nu = 1, \lambda_2 = 1$.

finite convergence material in Section 3.7.

In optimal control applications (see Chapter 5) it is advantageous to limit the number of times the contact function is evaluated. Thus in the selection rules which follow Y_{k+1}, $k \geq 0$, contains every point in Y_k except perhaps one. There remains the question of how to reject one contact point in favor of another. The approach in the selection rules given here is to use $\mu(z)$ as an indication of the quality of the contact point $s(-z)$, $z \in K$. Roughly speaking, contact points corresponding to larger values of $\mu(\cdot)$ are preferred. Other quantities, e.g., $|s(-z)|$, $|z|$, $\gamma(z)$, may be suggested for judging the merit of $s(z)$. However,



(a) BIP



(b) IIP Selection Rule A, $p = 2$.

Figure 3.5.2 Example 1 of Section 2.5, $\nu = 1$.

careful examination of example problems such as Example 3, Section 2.5 with $n = 3$, $\lambda_2 >> \lambda_3$ shows that these quantities are less desirable than $\mu(z)$.

It is convenient to view each selection rule as two phases, 1 and 2. Phase 1 is to be used for the first few iterations of IIP until a certain condition is satisfied, and Phase 2 is to be used thereafter.

Selection Rule A

Phase A1: For $k = 0$ set $y_i(0) = s(-z_0)$, $i = 1, 2, \ldots, p$, and define scalars $\mu_1, \mu_2, \ldots, \mu_p$ equal to $\mu(z_0)$. Whenever $0 < k \leq p$, set $y_i(k) = y_i(k-1)$, $i = 1, 2, \ldots, p$. Then set $y_k(k) = s(-z_{k-1})$ and $\mu_k = \mu(z_{k-1})$. When $k > p$ is satisfied, begin using Phase A2.

Phase A2: For each k set $y_i(k) = y_i(k-1)$, $i = 1, 2, \ldots, p$. Then let $\underline{\mu} = \min\{\mu_1, \mu_2, \ldots, \mu_p\}$ and let j be the smallest integer in $[1, p]$ for which $\mu_j = \underline{\mu}$. Whenever $\mu_j \leq \mu(z_{k-1})$, replace $y_j(k)$ by $s(-z_{k-1})$ and μ_j by $\mu(z_{k-1})$.

Selection Rule B

Phase B1: For $k = 0$ set $y_i(0) = s(-z_0)$, $i = 1, 2, \ldots, p$, and define scalars $\mu_1, \mu_2, \ldots, \mu_p$ equal to $\mu(z_0)$. When $k > 0$ is satisfied, begin using Phase B2.

Phase B2: Same as Phase A2.

COMMENT 3.5.1 In Phase A2(B2) there is nothing crucial about the way of handling the possibility of two or more μ_i, $1 \leq i \leq p$, being equal to $\underline{\mu}$. Moreover, $\mu_j < \mu(z_{k-1})$ may be used as the condition for replacement instead of $\mu_j \leq \mu(z_{k-1})$.

For selection rules such as A and B in which Y_{k+1} is a subset of the set $\{y_1(k), y_2(k), \ldots, y_p(k), s(-z_k)\}$, all $k \geq 0$, it is possible to state additional results. Let

$$S_k = \{y_1(k), y_2(k), \ldots, y_p(k), \; s(-z_k)\}, \quad k = 0, 1, 2, \ldots, \qquad (3.5.1)$$

and consider the determination of z_{k+1} in Step 2 of iteration k of IIP. Since $z_{k+1} \in H_k$, it has a representation: $z_{k+1} = \sum_{i=1}^{p+2} x^i y_i$ where $y_i = y_i(k)$ $(i = 1, 2, \ldots, p)$, $y_{p+1} = s(-z_k)$, $y_{p+2} = z_k$, $\sum_{i=1}^{p+2} x^i = 1$, $x^i \geq 0$ $(i = 1, 2, \ldots, p+2)$. Let I_k be the set of superscripts i, $1 \leq i \leq p + 2$, for which $x^i = 0$. Observe that I_k may be empty. The following results hold.

THEOREM 3.5.1 Consider the sequence $\{z_k\}$ generated by IIP. i) Suppose the selection rule is such that $Y_{k+1} \subset S_k$ for all $k \geq 0$. If in Step 2 of iteration k, $k > 0$, the integer $p + 1 \in I_k$, then $z_k = z_{k+1} = z^*$. ii) Suppose p is chosen so that $q = p + 2 - n \geq 0$. Then in Step 2 of iteration k, $k \geq 0$, either $z_{k+1} = 0$ (which implies $z_{k+1} = z^* = 0$) or there exists a vector $x = (x^1, x^2, \ldots, x^{p+2})$ such that the corresponding I_k contains at least q elements.

Now assume $z_{\bar{k}+1} \neq z^*$ for some $\bar{k} \geq 0$. Then for all k, $0 \leq k \leq \bar{k}$: iii) $Y_k \subset S_k \subset H_k \subset Q^-(z_{k+1}; z_{k+1})$; iv) if $Y_{k+1} \subset S_k$, $s(-z_{k+1}) \in Q^+(z_{k+1}; z_{k+1})$; v) if $Y_1 \subset S_0$, S_1 contains two distinct points; vi) if $Y_{k+1} \subset S_k$, Y_{k+1} contains every point in S_k except one, and S_k contains $\bar{p}$ distinct points $(1 \leq \bar{p} \leq p + 1)$, it follows that there are at least $\bar{p}$ distinct points in S_{k+1}; vii) if Phase A1 is used as the first phase in any selection rule and $z_p \neq z^*$, S_p contains $p + 1$ distinct points.

Proof: Consider i). Note that $k > 0$ and let $\bar{H} = \Delta\{y_1(k-1), y_2(k-1), \ldots, y_p(k-1), s(-z_{k-1}), z_k\}$. Since $\bar{H} \subset H_{k-1}$ and $z_k \in \bar{H}$, $|z_k| = \min_{z \in \bar{H}} |z|$. But $p + 1 \in I_k$ and $Y_k \subset S_{k-1}$ imply $z_{k+1} \in \bar{H}$ and $|z_{k+1}| \geq \min_{z \in \bar{H}} |z|$. Thus by part ii) of Theorem 3.4.1 $|z_k| = |z_{k+1}|$ and $z_k = z_{k+1} = z^*$. In ii) suppose $z_{k+1} \neq 0$. Part iii) of Theorem 3.2.1 yields $z_{k+1} \in \partial H_k$. Then part v) of the same theorem and $p + 2 - n \geq 0$ imply that a vector x with the desired property exists. Consider iii).

Since $z_{\bar{k}+1} \neq z^*$, part ii) of Theorem 3.4.1 implies $|z_k| > |z^*| \geq 0$ for all k, $0 \leq k \leq \bar{k}+1$. Hence, for $0 \leq k \leq \bar{k}$, the hyperplane $Q(z_{k+1};z_{k+1})$ is defined. For $w \neq 0$ let $P_{H_k}(w)$ be the support hyperplane of H_k with outward normal w. By part iv) of Theorem 3.2.1 it follows that $z_{k+1} \in P_{H_k}(-z_{k+1}) = Q(z_{k+1};z_{k+1})$. Thus H_k is contained in the closed half-space $Q^-(z_{k+1};z_{k+1})$ and iii) is true. In iv) suppose for some k, $0 \leq k \leq \bar{k}$, $s(-z_{k+1}) \notin Q^+(z_{k+1};z_{k+1})$. Then $s(-z_{k+1}) \in Q^-(z_{k+1};z_{k+1})$, and by iii) $Q^-(z_{k+1};z_{k+1})$ also contains S_k and z_{k+1}. Since $Y_{k+1} \subset S_k$, $H_{k+1} \subset Q^-(z_{k+1};z_{k+1})$ and $|z_{k+1}| = |z_{k+2}|$. By part ii) of Theorem 3.4.1 this implies $z_{k+1} = z^*$ for some k, $0 \leq k \leq \bar{k}$, which contradicts $|z_j| > |z^*|$, $0 \leq j \leq \bar{k}+1$, and thus established iv). From iii), iv) and $Y_1 \subset S_0$ it follows that $Y_1 \in Q^-(z_1;z_1)$ and $s(-z_1) \in Q^+(z_1;z_1)$. Hence S_1 must contain two distinct points. Similarly iii), iv), and $Y_{k+1} \subset S_k$ imply $Y_{k+1} \in Q^-(z_{k+1};z_{k+1})$ and $s(-z_{k+1}) \in Q^+(z_{k+1};z_{k+1})$, which means $s(-z_{k+1})$ is distinct from the points in Y_{k+1}. If Y_{k+1} contains every point in S_k except one and S_k contains $\bar{p}$ distinct points, then there are at least $\bar{p} - 1$ distinct points in Y_{k+1}. Since S_{k+1} is the union of Y_{k+1} and $s(-z_{k+1})$, the conclusion in vi) is true. Consider vii). From $z_p \neq z^*$, it follows that iii) and iv) hold for $0 \leq k \leq p-1$. Thus $Y_{k+1} \subset S_k$ and $0 \leq k \leq p-1$ imply $s(-z_{k+1})$ is distinct from the points in Y_{k+1}. For Phase A1 Y_1 and S_0 contain only $s(-z_0)$, Y_2 and S_1 contain only $s(-z_0)$ and $s(-z_1)$, ..., Y_p and S_{p-1} contain only $s(-z_0), s(-z_1), \ldots, s(-z_{p-1})$. Consequently the fact that $s(-z_{k+1}) \notin Y_{k+1}$ applied successively for $k = 0, 1, \ldots, p-1$ yields the result: $s(-z_0), s(-z_1), \ldots, s(-z_p)$ are distinct. Since S_p is the union of these $p+1$ points, vii) holds and the proof is complete.

COMMENT 3.5.2 Observe that if either Selection Rule A or Selection Rule B is used in IIP, $Y_{k+1} \subset S_k$ and Y_{k+1} contains every point in S_k except one for all $k \geq 0$. However, Rule A is generally better than Rule B because only p iterations are required with A to ensure that S_k contains $p+1$ distinct points (assuming $z_p \neq z^*$).

Now consider two additional selection rules.

Selection Rule C (Assume that $p \geq n$ and that in Step 2 of every iteration k of IIP a vector x is determined such that the corresponding I_k contains at least $p + 2 - n$ integers. By part ii) of Theorem 3.5.1 such an x must exist or $z_{k+1} = 0$. Terminate IIP in Step 2 of iteration k if $z_{k+1} = 0$ (which implies $z_{k+1} = z^* = 0$) or if $p + 1 \in I_k$ (which implies $z_k = z_{k+1} = z^*$)).

Phase C1: For $k = 0$ set $y_i(0) = s(-z_0)$, $i = 1, 2, \ldots, p$, and define scalars $\mu_1, \mu_2, \ldots, \mu_p$ equal to $\mu(z_0)$. When $k > 0$ and in Step 2 of iteration $k - 1$ the condition $p + 2 \in I_{k-1}$ is satisfied, begin using Phase C2. For $0 < k \leq p$, set $y_i(k) = y_i(k - 1)$, $i = 1, 2, \ldots, p$. Then set $y_k(k) = s(-z_{k-1})$ and $\mu_k = \mu(z_{k-1})$. For $k > p$, set $y_i(k) = y_i(k - 1)$, $i = 1, 2, \ldots, p$. Then let $\underline{\mu} = \min\{\mu_1, \mu_2, \ldots, \mu_p\}$ and let j be the smallest integer in $[1, p]$ for which $\mu_j = \underline{\mu}$. Whenever $\mu_j \leq \mu(z_{k-1})$, replace $y_j(k)$ by $s(-z_{k-1})$ and μ_j by $\mu(z_{k-1})$.

Phase C2: For $0 < k \leq p$, set $y_i(k) = y_i(k - 1)$, $i = 1, 2, \ldots, p$. Then set $y_k(k) = s(-z_{k-1})$ and $\mu_k = \mu(z_{k-1})$. For $k > p$, set $y_i(k) = y_i(k - 1)$, $i = 1, 2, \ldots, p$. Then let $\underline{\mu}' = \min_{\substack{i \in I_{k-1} \\ i \neq p+1,\, p+2}} \mu_i$ and let j be the smallest integer in $\{i : i \in I_{k-1},\ i \neq p + 1,\ p + 2\}$ for which $\mu_j = \underline{\mu}'$. Replace $y_j(k)$ by $s(-z_{k-1})$ and μ_j by $\mu(z_{k-1})$.

COMMENT 3.5.3 For a particular problem it is possible that the condition $p + 2 \in I_{k-1}$, $k > 0$, for entering Phase C2 may never be satisfied. In that case Selection Rules A and C are identical. The computational experience with IIP indicates, however, that this condition is satisfied after only a few iterations for a broad class of problems.

Selection Rule D (Assume that $p \geq n$, that $z_0 = s(-z_{-1})$ for some $z_{-1} \in K$, and that in Step 2 of every iteration k of IIP a vector x is determined such that the corresponding I_k contains at least $p + 2 - n$ integers. By part ii) of Theorem 3.5.1 such an x must exist or $z_{k+1} = 0$. Terminate IIP in Step 2 of iteration k if $z_{k+1} = 0$ (which implies $z_{k+1} = z^* = 0$) or if $p + 1 \in I_k$ (which implies $z_k = z_{k+1} = z^*$)).

Phase D 1: For $k = 0$ set $y_1(0) = z_0 = s(-z_{-1})$, set $y_i(0) = s(-z_0)$, $i = 2, 3, \ldots, p$, and define $\mu_1 = \mu(z_{-1})$, $\mu_i = \mu(z_0)$, $i = 2, 3, \ldots, p$. When $k > 0$ is satisfied, begin using Phase D2.

Phase D2: For $0 < k < p$, set $y_i(k) = y_i(k-1)$, $i = 1, 2, \ldots, p$. Then set $y_{k+1}(k) = s(-z_{k-1})$ and $\mu_{k+1} = \mu(z_{k-1})$. For $k \geq p$, set $y_i(k) = y_i(k-1)$, $i = 1, 2, \ldots, p$. Then let $\underline{\mu}' = \min\limits_{\substack{i \in I_{k-1} \\ i \neq p+1,\, p+2}} \mu_i$ and let j be the smallest integer in $\{i : i \in I_{k-1},\ i \neq p+1,\ p+2\}$ for which $\mu_j = \underline{\mu}'$. Replace $y_j(k)$ by $s(-z_{k-1})$ and μ_j by $\mu(z_{k-1})$.

COMMENT 3.5.4 The assumption $p \geq n$ is required for Selection Rule C and Selection Rule D so that the set $\{i : i \in I_{k-1},\ i \neq p+1,\ p+2\}$ which occurs in Phases C2 and D2 is not empty. Since $p + 2 - n \geq 2$, I_{k-1} contains at least 2 integers in $[1, p+2]$. Moreover, $p + 1 \notin I_{k-1}$ or IIP would have terminated in Step 2 of iteration $k - 1$.

COMMENT 3.5.5 Observe that if either Selection Rule C or Selection Rule D is used in IIP, $Y_{k+1} \subset S_k$ and Y_{k+1} contains every point in S_k except one for all $k \geq 0$. Furthermore, arguments like those used for Selection Rule A show that if $z_p \neq z^*$, S_p (with Rule C) and S_{p-1} (with Rule D) contain $p + 1$ distinct points.

THEOREM 3.5.2 *Consider IIP and assume that Selection Rule C or Selection Rule D is used. Let* $\hat{k}$ *be the first* $k \geq 0$ *for which* $p + 2 \in I_k$ *if Selection Rule C is used and let* $\hat{k} = 0$ *if Selection Rule D is used. Then in Step 2 of iteration k, all* $k \geq \hat{k}$, *Subproblem 1 can be solved on* ΔS_k *instead of* H_k. *That is, let* z_{k+1} *satisfy* $z_{k+1} \in \Delta S_k$, $|z_{k+1}| = \min_{z \in \Delta S_k} |z|$ *and find a vector* $x = (x^1, x^2, \ldots, x^{p+2})$ *such that* $z_{k+1} = \sum_{i=1}^{p+2} x^i y_i$ *where* $y_i = y_i(k)$ $(i = 1, 2, \ldots, p)$, $y_{p+1} = s(-z_k)$, $y_{p+2} = z_k$, $\sum_{i=1}^{p+2} x^i = 1$, $x^i \geq 0$ $(i = 1, 2, \ldots, p+1)$, $x^{p+2} = 0$.

Proof: Since $p + 2 \in I_{\hat{k}}$ (with Rule C) and $y_1(\hat{k}) = z_{\hat{k}}$ (with Rule D), on iteration $\hat{k}$ Subproblem 1 can certainly be solved on $\Delta S_{\hat{k}}$ instead of $H_{\hat{k}}$. Thus $z_{\hat{k}+1} \in \Delta S_{\hat{k}}$. By comment 3.5.5 $Y_{\hat{k}+1} \subset S_{\hat{k}}$. Furthermore, Selection Rules C and D are such that $Y_{\hat{k}+1}$ contains every point in $S_{\hat{k}}$ except one which has a coefficient of 0 in the convex combination expression for $z_{\hat{k}+1}$. Hence $z_{\hat{k}+1} \in \Delta Y_{\hat{k}+1}$ and $\Delta S_{\hat{k}+1} = H_{\hat{k}+1}$, so that on iteration $\hat{k} + 1$ Subproblem 1 can be solved on $\Delta S_{\hat{k}+1}$ to yield $z_{\hat{k}+2} \in \Delta S_{\hat{k}+1}$. By induction Subproblem 1 can be solved on ΔS_k instead of H_k for all iterations k, $k \geq \hat{k}$. This completes the proof.

Note that it is simpler to solve Subproblem 1 on ΔS_k rather than on H_k: the constraint set for the quadratic programming problem is the convex hull of only $p + 1$ points instead of $p + 2$. It will henceforth be assumed that whenever Selection Rule C or Selection Rule D is used in IIP, Subproblem 1 is solved on ΔS_k for all $k \geq \hat{k}$.

Section 3.8 contains computational results for IIP with Selection Rules A, B and C. These results indicate that it is good to choose $p = n$ and that for a broad class of problems, IIP (with $p = n$ and any of the Selection Rules A, B, C) exhibits much more rapid convergence than BIP. Selection Rule D is identical with Selection Rule C for $k > \max\{p, \tilde{k}\}$, where $\tilde{k}$ is the first $k \geq 0$ for which $p + 2 \in I_k$ when

Rule C is used. For all the computations in which Selection Rule C was used, $\tilde{k}$ was observed to be very small. Thus it can be stated that IIP with Selection Rule D also converges much more rapidly than BIP.

The computational results show, furthermore, that IIP converges at about the same rate with Selection Rule A or Selection Rule C. Thus the decision of which selection rule to use may be based on other considerations.

By comment 3.5.2 Selection Rule B may be rejected in favor of Selection Rule A. From Theorem 3.5.2 Selection Rules C and D have an advantage over Rule A in that for iterations k, $k \geq \hat{k}$, Subproblem 1 can be solved on the convex hull of $p + 1$ points instead of $p + 2$. However, the requirement with Rules C and D that I_k contain at least $p + 2 - n$ integers adds complexity to the solution of Subproblem 1 in Step 2 of every iteration k, $k \geq 0$ (see Section 3.6). Selection Rule D is most desirable for guaranteeing finite convergence in certain problems (see Section 3.7) and Rules C and D are advantageous for certain optimal control applications (see Chapter 5).

3.6 Solution of Subproblem 1

Since each iteration of IIP requires the solution of Subproblem 1, it is important that methods exist for readily computing its solution. As mentioned in Section 3.2 Subproblem 1 is a quadratic programming problem on a convex polyhedron constraint set. This is the type of problem that is usually described in the literature [e.g., A1, B2, H1, V1] under the heading "quadratic programming". However, the computational algorithms which are suggested always begin by assuming the constraint set is described by a set of linear equations and/or inequalities rather than by the points whose convex hull is the constraint polyhedron. Thus to apply the standard quadratic programming techniques directly to Subproblem 1 it is necessary to first determine from these points a description of the constraint set in terms of linear equations and/or inequalities. Such a determination presents nearly insurmountable computational difficulties.

There is an alternative method of attacking Subproblem 1 which makes possible the use of the standard algorithms. It is shown in this section that the solution to Subproblem 1 is given by the solution to another quadratic programming problem, Subproblem 2, which has a constraint set described by linear equations and inequalities. Subproblem 2 is solvable by any of the well-known quadratic programming techniques such as those due to Frank and Wolfe [F4], Wolfe [W1], Beale [B4, B5], Houthakker [H4], Hildreth [H2], Markowitz [M1], and Lemke [L2].

Let $y_1, y_2, \ldots, y_m \in E^n$ be the known points whose convex hull is the convex polyhedron $H = \Delta\{y_1, y_2, \ldots, y_m\}$ specified in Subproblem 1. Note that since $|\cdot|$ is the Euclidean norm, an equivalent statement of Subproblem 1 is: find $y^* \in H$ such that $|y^*|^2 = \min_{y \in H} |y|^2$.

Each $y \in H$ has the representation $y = \sum_{i=1}^{m} x^i y_i$ where $\sum_{i=1}^{m} x^i = 1$, $x^i \geq 0$ $(i = 1, 2, \ldots, m)$. Thus

$$|y|^2 = \left|\sum_{i=1}^{m} x^i y_i\right|^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} x^i x^j y_i \cdot y_j \; . \tag{3.6.1}$$

If x is the m-vector $(x^1, x^2, \ldots, x^m)$ and D is the $m \times m$ symmetric matrix with elements $d_{ij} = y_i \cdot y_j$, then

$$|y|^2 = x \cdot Dx \; . \tag{3.6.2}$$

Since $|y|^2 \geq 0$, the quadratic form $x \cdot Dx$ is non-negative definite, a fact which implies it is a convex function of x on E^m. Consider now the following quadratic programming problem.

SUBPROBLEM 2 Given: D, an $m \times m$ symmetric non-negative definite matrix, and the constraint set

$$X = \{x \in E^m : \sum_{i=1}^{m} x^i = 1,\ x^i \geq 0\ (i = 1, 2, \ldots, m)\} .$$

<u>Find</u>: <u>a</u> <u>point</u> $x^* \in X$ <u>such</u> <u>that</u> $x^* \cdot Dx^* = \min_{x \in X} x \cdot Dx$.

If $D = [d_{ij}] = [y_i \cdot y_j]$, Subproblem 2 is said to be associated with Subproblem 1. For this case (3.6.2) implies that minimization of $x \cdot Dx$ on X is equivalent to minimization of $|y|^2$ on H. Thus, if x^* solves Subproblem 2, y^* is given by

$$y^* = \sum_{i=1}^{m} x^{*i} y_i . \tag{3.6.3}$$

A solution $x^* \in E^m$ to Subproblem 2 may have as many as m non-zero elements. However, if Selection Rule C or Selection Rule D is used in Step 1 of IIP, it is required that a solution to the Subproblem 2 associated with Subproblem 1 in Step 2 of IIP be obtained which has at least m - n zero elements. If no such solution exists, then IIP terminates and y^*, the solution to Subproblem 1, and z^*, the solution to BP, both equal 0.

Assume $m \geq n$ and consider Subproblem 1 and the associated Subproblem 2. Suppose a solution x_1^* to Subproblem 2 is obtained for which more than n elements, say $\tilde{n}$, are nonzero. It will now be shown how a second solution x_2^* to Subproblem 2 which has at least m - n zero elements can be determined, provided it exists.

The point y^* is given by $y^* = \sum_{i=1}^{m} x_1^{*i} y_i$. Thus $y^* = \sum_{i=1}^{m} x_2^{*i} y_i$ represents a system of n equations in m unknowns: $x_2^{*1}, x_2^{*2}, \ldots, x_2^{*m}$. The $n \times m$ matrix of coefficients $[y_1, y_2 \cdots y_m]$ has rank $\leq n$ so that at least m - n of the unknowns x_2^{*i} may be given arbitrary values. There are $m - \tilde{n}$ zero elements of x_1^* so set the corresponding elements of x_2^* equal to zero. It is required to find $\tilde{n} - n$ of the remaining $\tilde{n}$ unknowns which can be set equal to zero.

This can be determined by trial-and-error where the number of trials is at most $\frac{\tilde{n}!}{(\tilde{n} - n)!\,n!}$. Suppose $\tilde{n} - n$ of the remaining unknowns are selected arbitrarily and set equal to zero. There results a set of n equations in n unknowns. These equations may present computational difficulties because a solution does not necessarily exist, and even if it exists it may not be unique. If there is a solution such that the resulting x_2^* satisfies $\sum_{i=1}^{m} x_2^{*i} = 1$, $x_2^{*i} \geq 0$ $(i = 1, 2, \ldots, m)$, then this is the desired x_2^*. Otherwise another set of $\tilde{n} - n$ unknowns are selected and the procedure is repeated.

If the solution x_1^* indicates that $y^* \neq 0$, there is another way of obtaining y^* in the form $y^* = \sum_{i=1}^{m} \sigma_i y_i$ where $\sum_{i=1}^{m} \sigma_i = 1$, $\sigma_i \geq 0$ $(i = 1, 2, \ldots, m)$, and at least $m - n$ of the σ_i, $1 \leq i \leq m$, are zero. Parts iii) and v) of Theorem 3.2.1 and $y^* \neq 0$ imply that this form exists and that $y^* \in \partial H$. The method is to consider the $\bar{m} = \frac{m!}{(m-n)!\,n!}$ convex polyhedra $\Delta_1, \Delta_2, \ldots, \Delta_{\bar{m}}$ formed by the convex hull of n points chosen from $\{y_1, y_2, \ldots, y_m\}$. For each Δ_j $(j = 1, 2, \ldots, \bar{m})$ a Subproblem 1 and its associated Subproblem 2 (D an $n \times n$ matrix here) can be solved to yield a point $y_j^* \in \Delta_j$ such that $|y_j^*| = \min_{y \in \Delta_j} |y|$. If $y^*_{j\,min}$ $(1 \leq j\,min \leq \bar{m})$ satisfies $|y^*_{j\,min}| = \min_{1 \leq j \leq \bar{m}} |y_j^*|$, then part v) of Theorem 3.1.2 and $y^* \in \partial H$ imply that $y^* = y^*_{j\,m}$. Thus y^* can be written as a convex combination of n points from $\{y_1, y_2, \ldots, y_m\}$, which is the desired form.

Of the two methods just described the former is simpler because it requires a Subproblem 2 to be solved just once instead of $\bar{m} + 1$ times.

It may be possible to avoid entirely the possibility of trial-and-error or solution of subproblems on Δ_j $(j = 1, 2, \ldots, \bar{m})$. This is the case if a quadratic programming technique can be found for Subproblem 2 which yields a solution with a maximum number of zero elements. Considering the simple nature of the constraint set X it is likely that this would not be difficult to do.

3.7 A Finite Convergence Theorem

The following theorem gives a sufficient condition for IIP to exhibit finite convergence.

THEOREM 3.7.1 Let $s(\cdot)$ be an arbitrary contact function of the set K specified in BP, choose $p \geq n$, and consider IIP with Selection Rule D. Assume that K is a convex polyhedron and that the range of $s(y)$ for $y \in E^n$ is a finite set of points. Then the sequence $\{z_k\}$ generated by IIP converges in a finite number of iterations.

If K is a convex polygon in $E^n (n = 2)$, then the range of $s(y)$ for $y \in E^n$ is a finite set of points. This conclusion is not necessarily true, however, for a convex polyhedron K in E^n, $n \geq 3$. For example, in E^3 an entire edge of K could lie in the range of $s(y)$, $y \in E^3$. Nevertheless, for many convex polyhedra K such as those which arise in optimization problems for linear sampled-data systems, it may be possible to choose a contact function $s(y)$ of K whose range for $y \in E^n$ is a finite set of points.

Proof of Theorem 3.7.1: Consider first the following result.

LEMMA 3.7.1 If K is a convex polyhedron in E^n and $s(\cdot)$ is an arbitrary contact function of K, then every extreme point of K lies in the range of $s(y)$, $y \in E^n$.

Proof: Let x be any extreme point of K. It can be shown that there exists a support hyperplane of K that contacts K in the single point x. Therefore, if y is the outward normal to this support hyperplane, $x = s(y)$ and the lemma is proved.

Now let $\bar{s}_1, \bar{s}_2, \ldots, \bar{s}_\ell$ denote the points in the range of $s(y)$ $y \in E^n$ and define $\bar{S} = \{\bar{s}_1, \bar{s}_2, \ldots, \bar{s}_\ell\}$. By Lemma 3.7.1 every extreme point of $K \in \bar{S}$. This and $\bar{S} \subset K$ imply $K = \Delta\bar{S}$. Thus $z^* \in \Delta\bar{S}$.

Suppose that the convergence of IIP is not finite. From

Comment 3.5.5 and part vi) of Theorem 3.5.1, S_k, $k \geq p - 1$, contains $p + 1$ distinct contact points when Selection Rule D is used. Each of these points is in $\bar{S}$. If $\ell < p + 1$, a contradiction has already been obtained so henceforth consider $\ell \geq p + 1$. Since $p \geq n$, this implies $\ell \geq n + 1$.

Let $\Delta_1, \Delta_2, \ldots, \Delta_{\bar{\ell}}$, $\bar{\ell} = \dfrac{\ell!}{(n+1)!(\ell - n - 1)!}$, denote the convex polyhedra formed by the convex hull of $n + 1$ points chosen from $\bar{S}$. By part vii) of Theorem 3.1.2, $\Delta\bar{S} = \bigcup_{j=1}^{\bar{\ell}} \Delta_j$. Thus $z^* \in \Delta\bar{S}$ implies $z^* \in \Delta_j$ for at least one j, $i \leq j \leq \bar{\ell}$. Let $\delta_j (j = 1, 2, \ldots, \bar{\ell})$ be the distance of each set Δ_j from z^*; that is, $\delta_j = \min_{z \in \Delta_j} |z - z^*|$. At least one δ_j, $1 \leq j \leq \bar{\ell}$, is 0 and $\delta_j = 0$ if and only if $z^* \in \Delta_j$

Since S_k, $k \geq p - 1$, contains $p + 1$ distinct points from $\bar{S}$ and $p \geq n$, there is at least one Δ_j, $1 \leq j \leq \bar{\ell}$, such that $\Delta_j \subset \Delta S_k$. Furthermore, part vii) of Theorem 3.1.2 implies $\bigcup_{\substack{\Delta_j \subset \Delta S_k \\ 1 \leq j \leq \bar{\ell}}} \Delta_j = \Delta S_k$. Let $\delta(k)$ be defined for $k \geq p - 1$ by: $\delta(k) = \min_{\substack{\Delta_j \subset \Delta S_k \\ 1 \leq j \leq \bar{\ell}}} \delta_j$. That is, $\delta(k) = \min_{z \in \Delta S_k} |z - z^*|$. By Theorem 3.5.2 it follows that $z_k \in \Delta S_k = H_k$, $k \geq 0$, when Selection Rule D is used. Hence $|z_k - z^*| \geq \delta(k) \geq 0$, all $k \geq p - 1$. Part iii) of Theorem 3.4.1 implies $|z_k - z^*| \to 0$ as $k \to \infty$, so $\delta(k) \to 0$ as $k \to \infty$. But $\delta(k)$ has $\bar{\ell}$ values at most. Consequently there exists $\bar{k} \geq p - 1$ for which $\delta(\bar{k}) = 0$. This implies $z^* \in \Delta S_{\bar{k}}$ and thus $z_{\bar{k}+1} = z^*$, contradicting the supposition that convergence is not finite. This completes the proof.

COMMENT 3.7.1 If there exists $k \geq 0$ such that the condition $p + 2 \in I_k$ is satisfied in Step 2 of iteration k, then Theorem 3.7.1 holds for IIP with Selection Rule C.

COMMENT 3.7.2 If $0 \notin K$, the assumption in Theorem 3.7.1 may be replaced by the following two less restrictive assumptions:

i) there exists $\epsilon > 0$ such that for $z \in K$ and $|z - z^*| < \epsilon$, $s(-z) \in S(-z^*)$;
ii) $S(-z^*)$ is a convex polyhedron and there are only a finite number of contact points contained in $S(-z^*)$ which lie in the range of $s(-z)$, $z \in K$.
Then the sequence $\{z_k\}$ generated by IIP with Selection Rule D converges in a finite number of iterations.

The proof of this result is based on the fact that for $0 \notin K$, $\mu(z_k) \to |z^*|$. By supposing convergence is not finite it can be shown that there exists k' such that for $k \geq k'$: $S_k \subset S(-z^*)$, $z_k \in S(-z^*)$, and S_k contains p + 1 distinct points. Then in much the same manner as in the proof of Theorem 3.7.1 a contradiction can be established.

Figure 3.7.1 illustrates a set K in E^2 which satisfies assumption ii) but not i). By inspection it is clear that unless $z_0 \in K$ is such that $z_0 = \omega z^*$, $\omega > 0$, convergence is not finite.

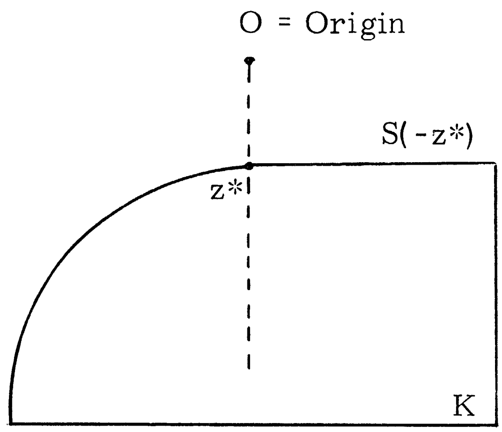


Figure 3.7.1 A $K \subset E^2$ which does not satisfy assumption i) of Comment 3.7.2.

3.8 Numerical Results for IIP

This section contains results of extensive computations with IIP for Example 3 of Section 2.5. As in Section 2.6 $\nu = 1$, $z^* = (1, 0, \ldots, 0)$, and the extent of K is increased beyond $z^1 = 2\nu$ so that (2.5.18) is valid. Data is presented for $n = 2, 3, \ldots, 6$; various p, $\lambda_2, \lambda_3, \ldots, \lambda_6$, z_0; and Selection Rules A, B, and C.

The quadratic programming technique of Frank and Wolfe [F4] was used to solve the Subproblem 2 which occurs on each iteration of IIP. This technique, which exhibits finite convergence, makes possible the use of the simplex method of linear programming [D1] as a subroutine. Briefly, the technique is as follows.

If D is the $m \times m$ matrix in Subproblem 2, define the $(m+1) \times (2m+2)$ matrix $\overline{D}$ by

$$\overline{D} = \begin{bmatrix} 1 & 1 \ldots 1 & 0 & 0 & 0 & 0 \ldots 0 \\ & & -1 & 1 & & \\ & & -1 & 1 & & \\ & -2D & \vdots & \vdots & I & \\ & & -1 & 1 & & \end{bmatrix} \qquad (3.8.1)$$

Furthermore, if r is the column vector (x, ζ, ξ, v), where $x, v \in E^m$ and $\zeta, \xi \in E^1$, let r^+ denote the row vector $(v, 0, 0, x)$. Note that $r^+r = 2x \cdot v$ and consider

SUBPROBLEM 3 <u>Given</u>: $\overline{D}$ <u>as in</u> (3.8.1). <u>Find</u>: $r \in E^{2m+2}$ <u>such that</u> $\overline{D}r = (1, 0, \ldots, 0)$, $r \geq 0$, <u>and</u> r^+r <u>is a minimum</u>.

By employing the Kuhn-Tucker conditions [K3] Frank and Wolfe show that if a solution x^* to Subproblem 2 exists then there is a solution r to Subproblem 3 for which $r^+r = 0$. Moreover, if such an r is found, its first m components are a solution to Subproblem 2.

Subproblem 3 is solved in the following two phases (a <u>feasible</u> vector $r \in E^{2m+2}$ satisfies the constraints $\overline{D}r = (1, 0, \ldots, 0)$, $r \geq 0$; a basic vector $r \in E^{2m+2}$ contains no more than m + 1 nonzero elements):

Phase I A basic feasible vector r_0 is found with which to begin Phase II (the simplex method with one artificial vector [H1] is easy to use here).

Phase II There are a finite number of iterations in this phase. On the h^{th} iteration a feasible vector w_h and a basic feasible vector r_{h-1} are available (on the first iteration use $w_1 = r_0$). Employ the simplex method to minimize the linear form $w_h^+ r$, obtaining the sequence of basic feasible vectors $\{r^j\}$, $j = 0, 1, 2, \ldots$, $r^0 = r_{h-1}$ such that $w_h^+ r^0 \geq w_h^+ r^1 > w_h^+ r^2 > \cdots$. Stop at the first r^j such that either $(r^j)^+ r^j = 0$ or $w_h^+ r^j \leq \frac{1}{2} w_h^+ w_h$. If the first condition is satisfied, r^j solves Subproblem 3. Otherwise let $r_h = r^j$, $\bar{\alpha}_h = \min\{\frac{w_h^+ (w_h - r_h)}{(r_h^+ - w_h^+)(r_h - w_h)}, 1\}$, and

$w_{h+1} = w_h + \bar{\alpha}_h (r_h - w_h)$; then repeat Phase II using w_{h+1} and r_h.

It is interesting to note that in all the computations with IIP, every solution x* to Subproblem 2 obtained by Frank and Wolfe's technique had at least m - n zero elements. Thus the difficulties (trial-and-error or solution of additional subproblems) mentioned in the latter part of Section 3.6 were not encountered.

Tables 3.8.1 through 3.8.9 (for IIP Selection Rule A and p = n) correspond directly to the nine tables in Section 2.6 (for BIP). Furthermore, Figures 3.8.1, 3.8.3, and 3.8.4 are similar to Figures 2.6.1, 2.6.3, and 2.6.4. These tables and figures show the marked improvement of IIP over BIP. This improvement has three facets: (1) much more rapid convergence, (2) very little dependence on the parameter $\bar{\lambda}\nu^{-1}$, and (3) no noticeable influence of the initial point z_0.

The behavior of $|z_k|-|z^*|$, $|z_k - z^*|$, $|z^*|-|z_k|\gamma(z_k)$, and $|s(-z_k) - z^*|$ shown in Figures 3.8.3 and 3.8.4 is typical of that observed in all computations with IIP. As with BIP it can be stated that $|z^*|-|z_k|\gamma(z_k)$ decreases most rapidly, followed in order by $|z_k|-|z^*|$, $|z_k - z^*|$, and $|s(-z_k) - z^*|$. Some other quantities related to the sequence $\{z_k\}$ are displayed in Tables 3.8.10, 3.8.11, and 3.8.12 (for $n = p = 3$ and IIP Selection Rule A), and Table 3.8.13 gives more data on $|z_k|-|z^*|$ for different λ_2/λ_3 ratios. As with BIP the parameter λ_2/λ_3 has very little effect.

It is clear from Table 3.8.14 that Selection Rule A is superior to Selection Rule B. Note that with Selection Rule B it may happen that a large number of iterations are required to satisfy $|z_k|-|z^*| \leq 1$, after which convergence is quite rapid. This shows the significance of S_k, $k \geq p$, containing $p + 1$ distinct points (see Comment 3.5.2).

Table 3.8.15 differs from Table 3.8.1 in that p equals 1 instead of 2. Tables 3.8.16 (for $n = 2$) and 3.8.17 (for $n = 3$) compare results for a variety of p values. This data and Table 3.8.20 (for $n = 4$) show that convergence is good for $p = n$ and little improvement is obtained for $p > n$. The desirability of choosing $p = n$ is also indicated by the actual computing time required to satisfy $|z_k| - |z^*| \leq 10^{-6}$. However, as mentioned in Section 2.6 the evaluation of a contact function in optimal control applications is the most time-consuming part of the iterative procedure. Thus the number of iterations to satisfy given error criteria is a better measure of the performance of the procedure.

Results for $n = p = 4, 5, 6$ and IIP Selection Rule A are given in Tables 3.8.18, 3.8.19, 3.8.21, 3.8.22, and 3.8.23. Roughly speaking, the rate of decrease of $|z_k|-|z^*|$ is dependent on n alone. The number of iterations per decade after a few initial iterations, is approximately 2 for $n = 2$, 4 for $n = 3$, 6 for $n = 4$, 9 for $n = 5$, and 13 for $n = 6$.

Finally, Tables 3.8.24 and 3.8.25 give results for IIP Selection Rule C. Observe that the rate of convergence in all cases is about the same as that for IIP Selection Rule A. The sequences $\{|z_k - z^*|\}$, $\{|z^*| - |z_k|\gamma(z_k)\}$, $\{|s(-z_k) - z^*|\}$, as well as $\{|z_k| - |z^*|\}$ behave similarly for IIP Selection Rules A and C.

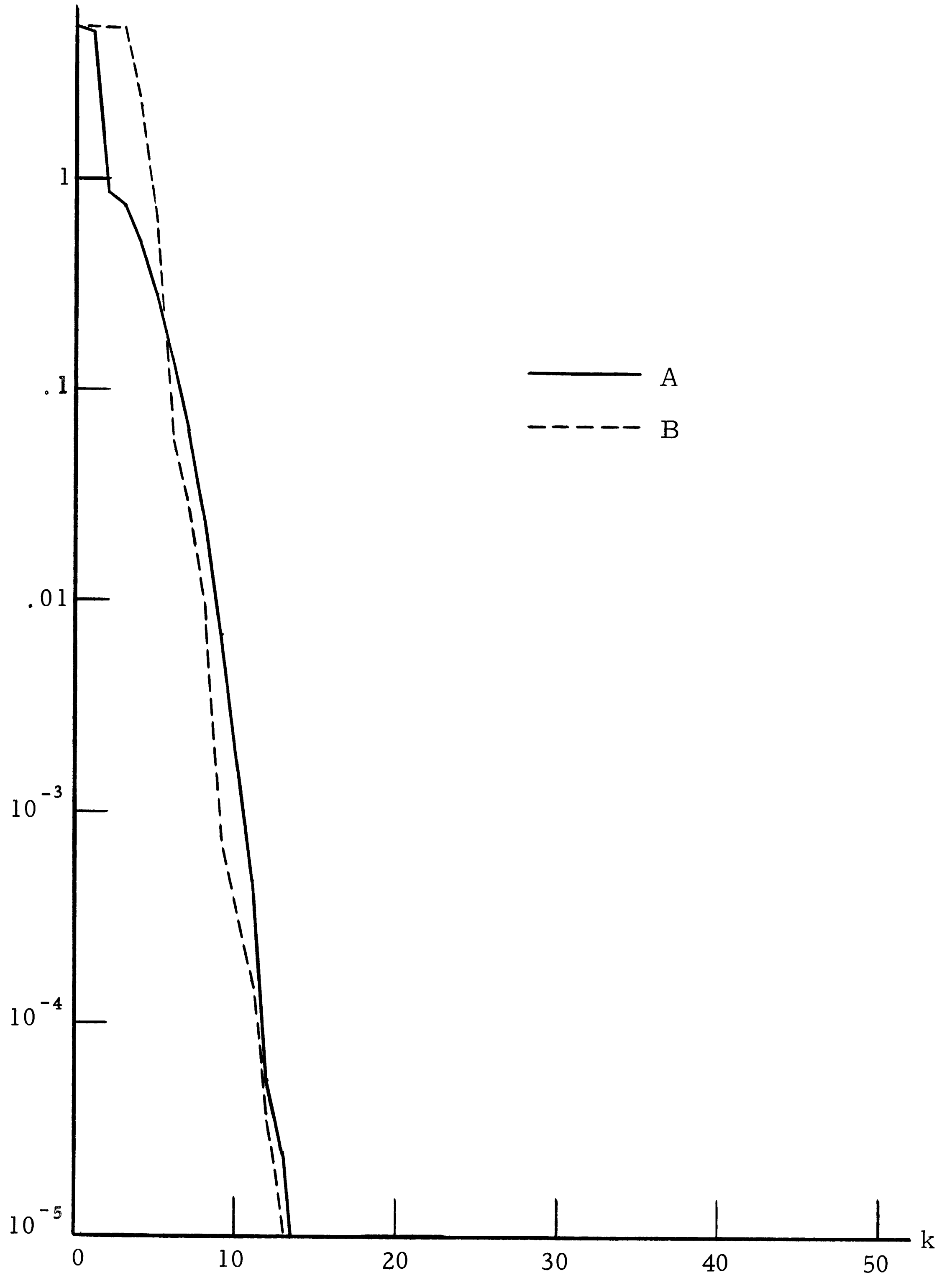


Figure 3.8.1 $|z_k| - |z^*|$ for $n = 2$, $p = 2$, $z_0 = (6, 2)$: A) $\lambda_2 = 100$, B) $\lambda_2 = 1000$; IIP Selection Rule A.

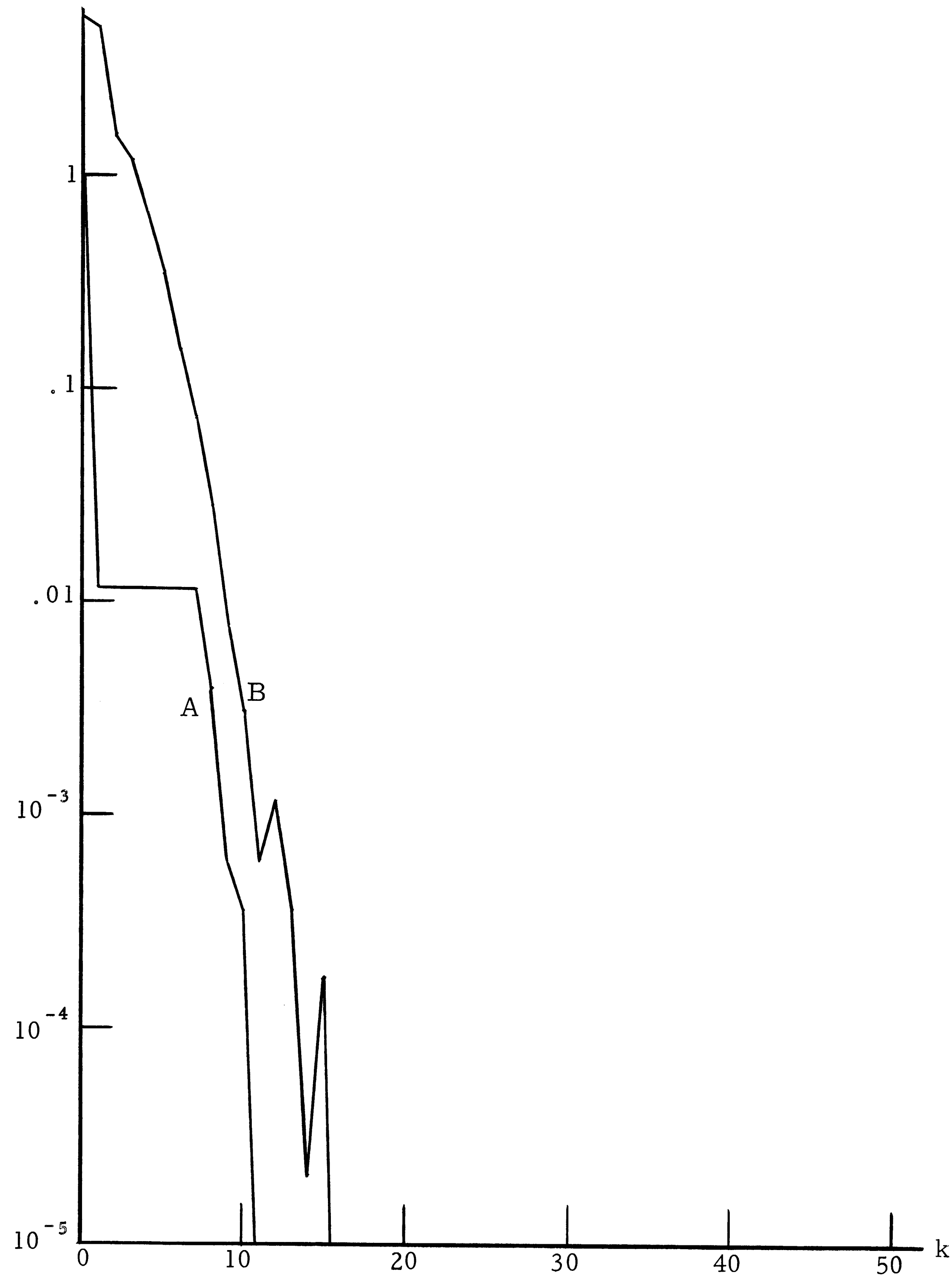


Figure 3.8.2 Results for $n = 2$, $p = 2$, $z_0 = (6, 2)$, $\lambda_2 = 100$: A) $|z^*| - \max_{i \le k} |z_i| \gamma(z_i)$, B) $|z_k - z^*|$; IIP Selection Rule A.

Table 3. 8. 1 Number of iterations to satisfy $|z_k| - |z^*| \le \epsilon$; $n = 2$, $p = 2$, $z_0 = (6, 2)$, IIP Selection Rule A.

λ_2 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10	3	4	6	7	10	11	13
100	2	7	9	11	12	14	15
200	4	5	7	9	10	10	12
500	3	5	7	7	10	12	14
1000	5	6	9	9	12	13	15
10^4	8	11	14	16	18	18	20
10^5	8	9	10	12	14	16	18
10^6	10	11	14	15	17	18	20

Table 3. 8. 2 Number of iterations to satisfy $|z_k - z^*| \le \epsilon$; $n = 2$, $p = 2$, $z_0 = (6, 2)$, IIP Selection Rule A.

λ_2 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10	3	5	6	9	12	>13	
100	4	7	9	11	14	>15	
200	4	5	8	9	>12		
500	3	5	7	9	11	>14	
1000	5	6	9	10	12	14	>15
10^4	8	11	14	16	18	>20	
10^5	8	9	10	12	14	16	18
10^6	10	11	14	15	17	18	>20

Table 3. 8. 3 Number of iterations to satisfy $|z^*| - |z_k|\gamma(z_k) \leq \epsilon$;

$n = 2$, $p = 2$, $z_0 = (6, 2)$, IIP Selection Rule A.

λ_2 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10	0	3	3	6	8	9	12
100	0	1	8	9	11	11	14
200	0	4	4	7	9	9	9
500	0	2	6	6	6	11	11
1000	0	5	5	8	8	12	14
10^4	0	4	4	15	17	17	17
10^5	0	7	9	9	13	15	15
10^6	0	10	10	13	16	16	19

Table 3. 8. 4 Number of iterations to satisfy $|s(-z_k) - z^*| \leq \epsilon$;

$n = 2$, $p = 2$, $z_0 = (6, 2)$, IIP Selection Rule A.

λ_2 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10	3	6	9	12	>13		
100	8	11	14	>15			
200	7	9	9	>12			
500	6	11	14	>14			
1000	8	12	14	>15			
10^4	17	17	20	>20			
10^5	15	18	>18				
10^6	19	>20					

Table 3. 8. 5 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$;
$n = 2$, $p = 2$, $\lambda_2 = 100$, IIP Selection Rule A.

z_0^1	z_0^2 \ ϵ	1	. 1	. 01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
6	1	2	5	9	12	13	14	15
6	1. 5	2	4	8	11	14	16	16
6	2	2	7	9	11	12	14	15
6	2. 5	3	6	7	10	11	12	13
6	3	2	7	10	12	14	15	17
6	3. 5	2	6	8	9	11	13	14
6	4	3	5	6	8	10	11	13
6	4. 5	4	8	10	11	12	14	15
6	5	3	7	8	10	12	13	14
2	6	2	6	9	12	12	13	15
3	5. 57	2	3	7	8	10	10	12
4	4. 9	3	7	9	10	12	14	16
4. 5	4. 44	2	6	8	11	13	16	18
5	3. 87	2	7	9	10	10	12	13
5. 5	3. 12	3	6	9	11	13	14	15
6. 2	1. 25	3	5	7	8	10	12	14

Note: For $z_0 = (6, 2)$ and the last seven cases in the table, $|z_0|^2 \cong 40$.

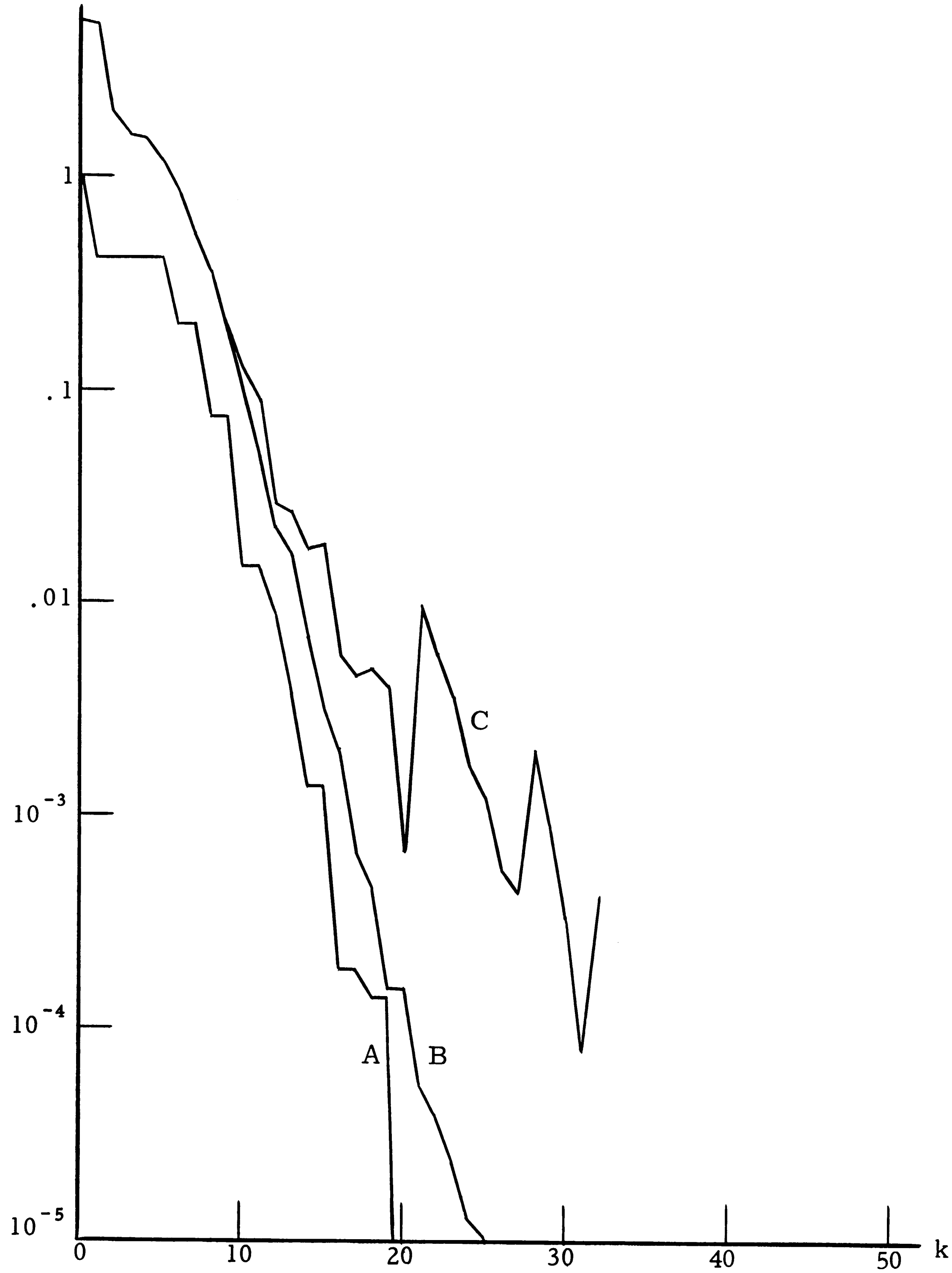


Figure 3.8.3 Results for $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, $\lambda_2 = 100$, $\lambda_3 = 10$: A) $|z^*| - \max_{i \le k} |z_i| \gamma(z_i)$, B) $|z_k| - |z^*|$, C) $|z_k - z^*|$; IIP Selection Rule A. For $k \le 9$, $|z_k - z^*| \cong |z_k| - |z^*|$.

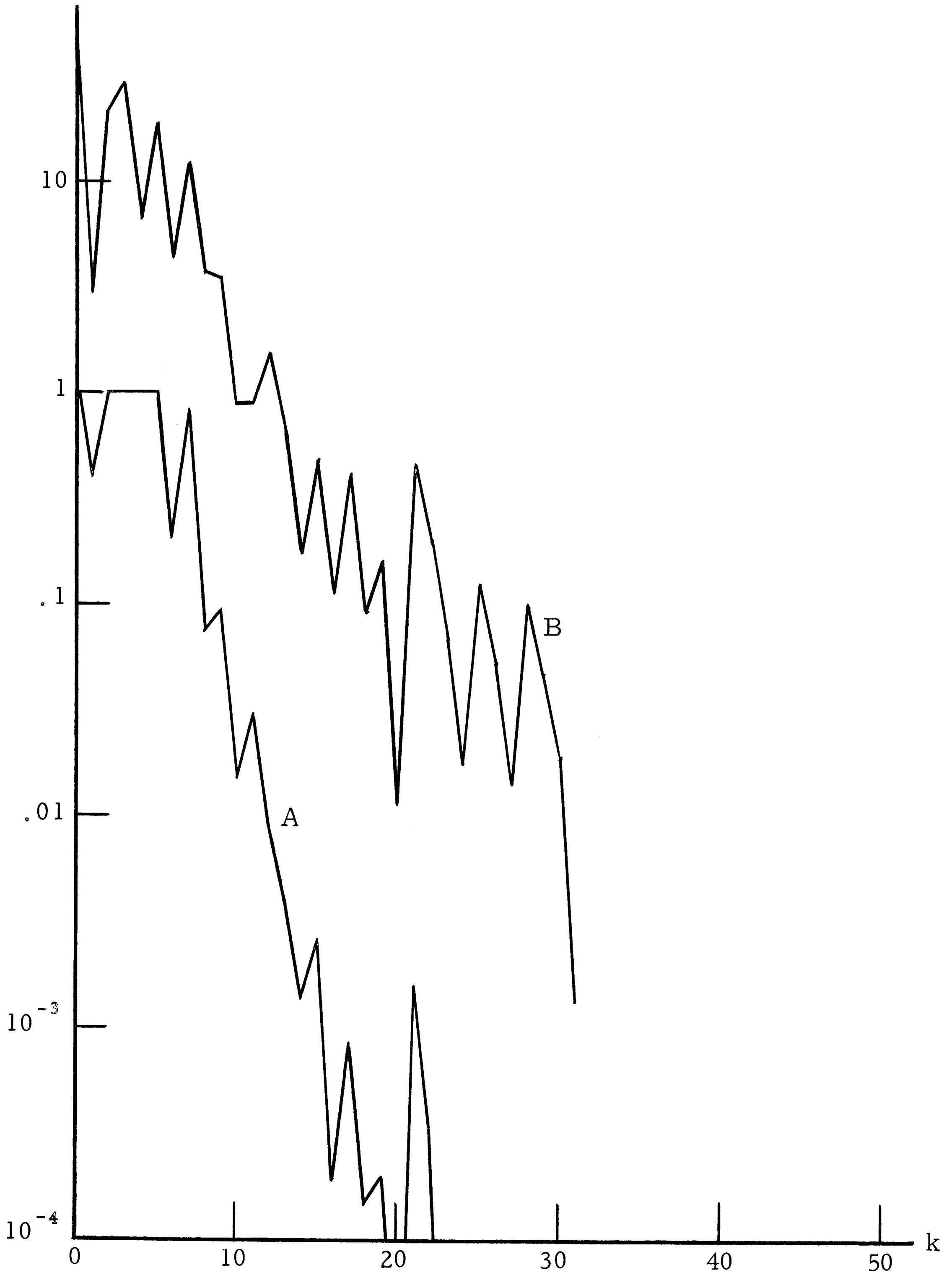


Figure 3.8.4 Results for $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, $\lambda_2 = 100$, $\lambda_3 = 10$: A) $|z^*| - |z_k|\gamma(z_k)$, B) $|s(-z_k) - z^*|$; IIP Selection Rule A.

Table 3.8.6 Number of iterations to satisfy $|z_k| - |z^*| \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	1	2	3	3	3	4	4
10	10	3	3	6	7	9	10	12
100	100	4	5	7	9	11	13	13
1000	1000	4	6	8	10	11	12	12
100	1	2	4	8	10	12	14	17
100	10	6	11	14	17	21	26	32
100	50	7	11	15	18	22	26	30
100	90	6	9	12	17	23	28	31
1000	1	6	7	8	11	13	15	17
1000	10	7	9	13	20	23	26	29
1000	100	8	14	17	23	26	29	32

Table 3.8.7 Number of iterations to satisfy $|z_k - z^*| \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	1	2	3	4	4	>4	
10	10	3	5	6	8	>12		
100	100	4	5	8	10	12	>13	
1000	1000	4	6	8	10	11	>12	
100	1	2	6	10	16	>17		
100	10	6	11	16	20	31	>32	
100	50	7	11	15	21	27	>30	
100	90	6	9	13	23	31	>31	
1000	1	6	7	12	12	16	>17	
1000	10	7	10	17	24	28	>29	
1000	100	8	14	18	25	30	>32	

Table 3.8.8 Number of iterations to satisfy $|z^*| - |z_k| \gamma(z_k) \leq \epsilon$;
$n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	0	2	2	2	2	3	3
10	10	0	2	2	6	8	8	11
100	100	0	4	4	8	9	10	12
1000	1000	0	5	7	8	9	11	11
100	1	0	1	7	9	11	13	16
100	10	0	8	12	16	20	20	31
100	50	0	9	13	15	19	24	27
100	90	0	7	11	13	18	25	27
1000	1	0	5	7	10	12	14	16
1000	10	0	7	12	14	22	25	28
1000	100	0	12	16	16	25	27	30

Table 3.8.9 Number of iterations to satisfy $|s(-z_k) - z^*| \leq \epsilon$;
$n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	0	2	3	3	4	4	>4
10	10	2	6	8	>12			
100	100	4	10	12	>13			
1000	1000	9	11	11	11	>12		
100	1	1	11	13	>17			
100	10	10	18	31	>32			
100	50	13	19	27	>30			
100	90	11	18	>31				
1000	1	7	14	>17				
1000	10	12	22	>29				
1000	100	16	27	>32				

Table 3. 8. 10 Number of iterations to satisfy $|z_k| - |z_k|\gamma(z_k) \leq \epsilon$;
$n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	1	2	3	3	3	4	4
10	10	3	3	6	8	9	11	> 12
100	100	4	5	7	9	11	> 13	
1000	1000	4	6	8	10	11	12	> 12
100	1	2	6	8	10	12	14	17
100	10	7	11	14	17	21	27	32
100	50	8	11	15	19	22	27	30
100	90	6	9	13	18	24	28	> 31
1000	1	6	7	8	11	13	16	17
1000	10	7	9	14	20	24	27	29
1000	100	9	14	17	24	27	30	32

Table 3. 8. 11 Number of iterations to satisfy $\sqrt{1 - \gamma(z_k)}\,|z_k| \leq \epsilon$;
$n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	2	3	4	> 4			
10	10	3	6	9	> 12			
100	100	5	8	11	> 13			
1000	1000	5	8	11	> 12			
100	1	3	9	12	> 17			
100	10	8	14	24	> 32			
100	50	9	15	23	> 30			
100	90	7	13	25	> 31			
1000	1	6	9	14	> 17			
1000	10	8	14	24	> 29			
1000	100	10	17	27	> 32			

Table 3.8.12 Number of iterations to satisfy $|s(-z_k)| - |z^*| \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
1	1	1	2	3	4	4	4	>4
10	10	2	2	6	8	8	11	>12
100	100	4	8	9	10	12	12	>13
1000	1000	8	9	11	11	11	11	11
100	1	1	9	11	11	13	13	>17
100	10	10	14	16	20	20	31	31
100	50	12	15	19	24	27	>30	
100	90	9	13	18	25	27	>31	
1000	1	7	12	14	16	>17		
1000	10	12	12	22	22	28	>29	
1000	100	16	23	27	30	30	>32	

Table 3.8.13 Number of iterations to satisfy $|z_k| - |z^*| \le \epsilon$; $n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule A.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^4	1	5	8	12	14	16	17	20
10^4	10	8	11	14	19	21	26	31
10^4	100	13	17	19	23	31	34	40
10^4	1000	14	19	23	26	30	34	37
10^4	5000	13	18	21	30	33	36	40
10^4	9000	13	15	21	25	28	37	41
10^5	1	8	8	12	14	17	19	23
10^5	10	9	13	16	20	25	29	33
10^5	100	10	19	23	25	32	35	40
10^5	1000	12	15	18	22	27	33	38
10^5	10^4	15	17	23	28	32	33	38

Table 3.8.14 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$;
$n = 3$, $p = 3$, $z_0 = (6, 2, 2)$, IIP Selection Rule B.

λ_2	λ_3 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}	$\hat{k}$
1	1	1	2	3	3	3	4	4	4
10	10	3	3	6	7	9	10	12	12
100	100	4	5	7	9	11	13	13	13
1000	1000	6	9	10	12	14	14	15	12
100	1	2	4	8	10	12	14	17	17
100	10	6	11	14	17	21	26	32	32
100	50	8	11	16	20	23	27	30	30
100	90	9	12	19	26	30	33	36	31
1000	1	6	7	8	11	13	15	17	17
1000	10	10	12	17	20	23	27	30	29
1000	100	23	39	42	46	47	54	57	32
10^4	1	8	11	14	17	19	21	23	20
10^4	10	10	16	18	25	28	30	34	31
10^4	100	54	57	61	63	67	71	73	40
10^4	1000	26	29	33	37	41	44	51	37
10^4	5000	52	63	67	70	75	83	88	40
10^4	9000	155	160	164	173	177	180	185	41
10^5	1000	165	168	171	174	179	185	191	38
10^5	10^4	55	59	62	65	68	76	79	38

Note: $\hat{k}$ = the first k for which $|z_k| - |z^*| \leq 10^{-6}$ with IIP Selection Rule A.

Table 3.8.15 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$; $n = 2$, $p = 1$, $z_0 = (6, 2)$, IIP Selection Rule A.

λ_2 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10	3	5	8	12	16	19	22
100	2	7	12	15	18	23	27
200	5	9	13	16	21	24	27
500	6	9	12	16	19	23	27
1000	6	10	15	19	23	28	32
10^4	8	11	14	20	22	26	30
10^5	10	15	19	21	25	29	32
10^6	10	14	18	21	25	28	31

Table 3.8.16 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$; $n = 2$, $z_0 = (6, 2)$, IIP Selection Rule A and BIP.

λ_2	p \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}	t
100	0	2	9	25	42	60	60	75	25.5
	1	2	7	12	15	18	23	27	29.1
	2	2	7	9	11	12	14	15	17.2
	3	2	7	9	11	12	14	15	18.9
500	0	36	113	126	134	150	158	241	82.6
	1	6	9	12	16	19	23	27	28.5
	2	3	5	7	7	10	12	14	15.2
	3	3	5	8	10	11	13	15	18.4
	4	3	5	8	10	12	13	15	21.5

Note: 1) p = 0 corresponds to BIP.
2) t = actual computing time (seconds) for IBM 7090.

Table 3.8.17 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$;
$n = 3$, $z_0 = (5, 4, 2)$, $\lambda_2 = 1000$, $\lambda_3 = 100$,
IIP Selection Rule A and BIP.

p \ ε	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}	t
0	53	77	138	202	241	298	366	125.3
1	30	52	97	134	171	249	293	319.1
2	9	15	27	36	46	52	58	75.9
3	9	16	23	26	29	33	36	50.1
4	8	14	17	24	27	32	37	77.8
5	8	13	16	23	26	31	35	90.3

Notes: 1) p = 0 corresponds to BIP.
2) t = actual computing time (seconds) for IBM 7090.

Table 3.8.18 Number of iterations to satisfy $|z_k| - |z^*| \le \epsilon$;
$n = 4$, $p = 4$, $z_0 = (6, 2, 2, 1)$, IIP Selection Rule A.

λ_2	λ_3	λ_4 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
100	10	1	5	9	15	19	23	27	29
100	10	10	6	10	15	20	25	28	33
100	50	10	7	14	23	29	36	43	49
100	50	50	7	10	13	17	20	23	27
100	90	10	7	13	19	28	34	41	48
100	90	30	9	14	21	28	37	42	49
100	90	50	8	14	19	23	31	37	43
100	90	70	9	17	23	30	35	43	52
100	100	50	6	9	13	16	20	22	26
1000	50	10	9	15	21	26	32	38	43
1000	100	1	7	9	14	18	22	25	29
1000	100	10	8	14	21	25	29	40	46
1000	500	100	10	15	26	31	38	42	48
1000	700	100	13	19	24	30	38	45	51
1000	900	500	13	19	30	41	47	57	63
1000	950	900	8	19	23	42	47	53	58
1000	995	990	4	15	21	31	44	49	54
10^4	5000	1000	18	28	35	44	48	59	64
10^5	5×10^4	10^4	16	23	33	39	46	52	57

Table 3.8.19 Number of iterations to satisfy $|z_k - z^*| \le \epsilon$;
$n = 4$, $p = 4$, $z_0 = (6, 2, 2, 1)$, IIP Selection Rule A.

λ_2	λ_3	λ_4 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
100	10	1	5	10	18	22	> 29		
100	10	10	6	10	19	27	32	> 33	
100	50	10	7	14	26	40	> 49		
100	50	50	7	10	16	19	> 27		
100	90	10	7	13	24	34	46	> 48	
100	90	30	9	15	25	36	> 49		
100	90	50	8	14	21	35	> 43		
100	90	70	9	17	26	37	51	> 52	
100	100	50	6	9	15	19	> 26		
1000	50	10	9	15	21	32	> 43		
1000	100	1	7	9	16	21	27	> 29	
1000	100	10	8	15	23	28	> 46		
1000	500	100	10	15	27	33	41	> 48	
1000	700	100	13	19	25	34	48	> 51	
1000	900	500	13	20	30	44	56	> 63	
1000	950	900	8	19	24	42	52	> 58	
1000	995	990	4	16	21	38	49	> 54	
10^4	5000	1000	18	28	35	44	51	> 64	
10^5	5×10^4	10^4	16	23	33	39	46	55	> 57

Table 3. 8. 20 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$;
$n = 4$, $z_0 = (6, 1, 2, 2)$, $\lambda_2 = 1000$, $\lambda_3 = 500$, $\lambda_4 = 100$,
IIP Selection Rule A.

p \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}	t
1	23	67	107	152	185	226	264	331.1
2	19	38	79	101	127	141	158	211.4
3	15	27	36	46	62	74	78	123.1
4	12	18	25	29	37	48	53	86.6
5	11	17	22	30	39	49	57	113.3

Note: t = actual computing time (seconds) for IBM 7090.

Table 3.8.21 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$;
$n = 5$, $p = 5$, $z_0 = (5, 3, 1, 1.8, 2.6)$,
IIP Selection Rule A.

λ_2	λ_3	λ_4	λ_5 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
100	50	30	10	8	19	30	39	42	54	64
100	70	50	10	9	20	29	39	48	58	66
100	90	50	10	10	20	31	39	47	58	67
1000	700	500	100	19	27	33	42	53	68	82
1000	900	500	100	17	25	34	43	51	72	79

Table 3.8.22 Number of iterations to satisfy $|z_k - z^*| \leq \epsilon$;
$n = 5$, $p = 5$, $z_0 = (5, 3, 1, 1.8, 2.6)$,
IIP Selection Rule A.

λ_2	λ_3	λ_4	λ_5 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
100	50	30	10	8	21	34	52	>64		
100	70	50	10	9	20	33	48	>66		
100	90	50	10	10	20	34	54	>67		
1000	700	500	100	20	29	39	62	80	>82	
1000	900	500	100	17	26	37	58	>79		

Table 3.8.23 Number of iterations to satisfy error criteria; $n = 6$, $p = 6$, $z_0 = (4, 3, 2.6, 2.6, 1.8, 1.8)$, IIP Selection Rule A.

Error Criterion	Λ \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
$\lvert z_k\rvert - \lvert z^*\rvert \le \epsilon$	Λ_1	10	24	34	53	64	79	93
	Λ_2	12	26	38	45	63	76	91
$\lvert z_k - z^*\rvert \le \epsilon$	Λ_1	10	24	41	73	>93		
	Λ_2	13	25	39	60	>91		
$\lvert z^*\rvert - \lvert z_k\rvert \gamma(z_k) \le \epsilon$	Λ_1	0	23	23	41	61	73	90
	Λ_2	0	20	31	39	60	74	85
$\lvert s(-z_k) - z^*\rvert \le \epsilon$	Λ_1	23	61	87	>93			
	Λ_2	33	60	90	>91			
$\lvert z_k\rvert - \lvert z_k\rvert \gamma(z_k) \le \epsilon$	Λ_1	10	24	37	56	67	82	>93
	Λ_2	14	26	39	47	65	79	>91
$\sqrt{1 - \gamma(z_k)}\,\lvert z_k\rvert \le \epsilon$	Λ_1	16	39	69	>93			
	Λ_2	16	39	68	>91			
$\lvert s(-z_k)\rvert - \lvert z^*\rvert \le \epsilon$	Λ_1	23	41	59	66	82	92	>93
		26	39	39	68	85	>91	

Note: Λ_1 : $\lambda_2 = 100$, $\lambda_3 = 70$, $\lambda_4 = 50$, $\lambda_5 = 30$, $\lambda_6 = 10$.
Λ_2 : $\lambda_2 = 100$, $\lambda_3 = 90$, $\lambda_4 = 70$, $\lambda_5 = 50$, $\lambda_6 = 10$.

Table 3.8.24 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$; IIP Selection Rule C.

n = p	λ_2	λ_3	λ_4	λ_5	λ_6 \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}	$\hat{k}$	k'
2	100					2	7	9	11	12	14	15	15	2
3	100	10				5	13	16	20	24	28	31	32	3
3	100	50				7	14	17	21	24	29	31	30	3
3	100	90				6	12	15	19	22	26	29	31	3
3	1000	100				9	15	18	23	27	30	32	32	7
4	100	50	10			8	15	25	30	38	45	51	49	3
4	100	90	10			7	14	21	28	34	42	47	48	3
4	100	90	50			9	18	24	31	35	43	48	43	3
4	100	90	70			7	13	25	33	38	47	51	52	3
4	1000	500	100			10	17	28	35	40	44	50	48	6
5	100	70	50	10		8	17	26	36	42	54	63	66	5
5	100	90	80	70		9	16	21	37	48	58	69	70	4
6	100	90	70	50	10	11	24	37	49	62	78	92	91	5
6	1000	90	70	50	10	15	29	45	60	73	84	94	95	7

Notes: 1) $\hat{k}$ = the first k for which $|z_k| - |z^*| \leq 10^{-6}$ with IIP Selection Rule A.

2) k' = the first k for which Phase C2 is used.

3) For n = 2, $z_0 = (6, 2)$; for n = 3, $z_0 = (6, 2, 2)$; for n = 4, $z_0 = (6, 2, 2, 1)$; for n = 5, $z_0 = (5, 3, 1, 1.8, 2.6)$; for n = 6, $z_0 = (4, 3, 2.6, 2.6, 1.8, 1.8)$.

Table 3.8.25 Number of iterations to satisfy $|z_k| - |z^*| \leq \epsilon$; IIP Selection Rule C.

n	p \ ϵ	1	.1	.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}	t
3	2	7	15	26	34	43	52	61	72.4
3	3	5	13	16	20	24	28	31	44.3
3	4	5	14	18	21	23	27	30	61.7
3	5	5	12	15	20	24	26	29	80.6
4	3	9	22	37	51	61	72	88	137.1
4	4	7	13	25	33	38	47	51	88.3
4	5	7	16	25	32	39	44	50	134.3
4	6	7	13	23	31	38	44	50	186.0
4	7	7	14	22	31	39	45	51	284.4

Notes: 1) t = actual computing time (seconds) for IBM 7090.
2) For n = 3: $z_0 = (6, 2, 2)$, $\lambda_2 = 100$, $\lambda_3 = 10$.
3) For n = 4: $z_0 = (6, 2, 2, 1)$, $\lambda_2 = 100$, $\lambda_3 = 90$, $\lambda_4 = 70$.

CHAPTER 4

THE GENERAL ITERATIVE PROCEDURE GIP

By following some of the ideas introduced by Fadden [Fl], it is possible to state a general problem which has application to a wide variety of optimal control problems. This general problem GP is formulated in this chapter and a method for solving it, called the general iterative procedure GIP, is described and shown to converge. Each iteration of GIP involves the minimization problem BP, which can be solved by IIP or BIP.

4.1 The General Problem GP

Let $\Omega = [0, \hat{\omega}]$, $\hat{\omega} > 0$, be a compact interval in E^1. For $\omega \in \Omega$ consider sets $K(\omega) \subset E^n$ which are compact and convex. Also let $K(\omega)$ be continuous on Ω, i. e., for every $\epsilon > 0$ there is a $\delta = \delta(\epsilon, \omega)$ such that $K(\omega) \subset K(\omega + \bar{\omega}) + N(0;\epsilon)$ and $K(\omega + \bar{\omega}) \subset K(\omega) + N(0;\epsilon)$ whenever $|\bar{\omega}| < \delta$ where ω, $\omega + \bar{\omega} \in \Omega$.

Before stating GP the support and contact functions for $K(\omega)$, $\omega \in \Omega$, are introduced. Let $\eta(\omega, y) = \max_{z \in K(\omega)} z \cdot y$ denote the <u>support function</u> of $K(\omega)$. Since $K(\omega)$ is compact, $\eta(\omega, y)$ is defined on $\Omega \times E^n$. As in Section 2.2, for fixed $\omega \in \Omega$ $\eta(\omega, \cdot)$ is a convex continuous function on E^n. The following result also holds.

THEOREM 4.1.1 <u>Given that the sets $K(\omega)$ are compact, convex, and continuous on Ω. The support function $\eta(\omega, y)$ is continuous as a function of ω and y on $\Omega \times E^n$.</u>

<u>Proof</u> (due to Fadden [Fl]): It is necessary to show that for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon, \omega, y)$ such that $|\eta(\bar{\omega}, \bar{y}) - \eta(\omega, y)| < \epsilon$ whenever $|(\bar{\omega}, \bar{y}) - (\omega, y)| < \delta$ where $\omega, \bar{\omega} \in \Omega$ and $y, \bar{y} \in E^n$. Now

$$|\eta(\bar{\omega}, \bar{y}) - \eta(\omega, y)| = |\eta(\bar{\omega}, \bar{y}) - \eta(\omega, \bar{y}) + \eta(\omega, \bar{y}) - \eta(\omega, y)|$$
$$\leq |\eta(\bar{\omega}, \bar{y}) - \eta(\omega, \bar{y})| + |\eta(\omega, \bar{y}) - \eta(\omega, y)|.$$

The continuity of $\eta(\omega, y)$ for fixed ω implies that given $\epsilon > 0$ there is a $\delta_1 = \delta_1(\epsilon, \omega, y)$ such that $|\eta(\omega, \bar{y}) - \eta(\omega, y)| < \frac{\epsilon}{2}$ whenever $|\bar{y} - y| < \delta_1$. That $\eta(\omega, y)$ is a continuous function of ω for fixed y follows from the continuity of $K(\omega)$. For every $\bar{\epsilon} > 0$ there is a $\bar{\delta} = \bar{\delta}(\bar{\epsilon}, \omega)$ such that $K(\omega) \subset K(\bar{\omega}) + N(0;\bar{\epsilon})$ and $K(\bar{\omega}) \subset K(\omega) + N(0;\bar{\epsilon})$ whenever $|\bar{\omega} - \omega| < \bar{\delta}$. Consider fixed $\bar{y}$ and note that the support function of $\overline{N}(0;\bar{\epsilon})$ is $\bar{\epsilon}|\bar{y}|$. Hence, $\eta(\omega, \bar{y}) < \eta(\bar{\omega}, \bar{y}) + \bar{\epsilon}|\bar{y}|$ and $\eta(\bar{\omega}, \bar{y}) < \eta(\omega, \bar{y}) + \bar{\epsilon}|\bar{y}|$. Thus $|\eta(\bar{\omega}, \bar{y}) - \eta(\omega, \bar{y})| < \bar{\epsilon}|\bar{y}|$ whenever $|\bar{\omega} - \omega| < \bar{\delta}$. Clearly for every $\epsilon > 0$ there exist a $\delta_2 = \delta_2(\epsilon, \omega, y)$ and a $\delta_3 > 0$ such that $|\eta(\bar{\omega}, \bar{y}) - \eta(\omega, \bar{y})| < \frac{\epsilon}{2}$ for all $\bar{y} \in N(y;\delta_3)$ whenever $|\bar{\omega} - \omega| < \delta_2$. Therefore, since $|(\bar{\omega}, \bar{y}) - (\omega, y)| \leq |\bar{\omega} - \omega| + |\bar{y} - y|$, the desired property is established if $\delta = \delta_2 + \min\{\delta_1, \delta_3\}$.

For $\omega \in \Omega$ let $P(\omega, y)$, $y \neq 0$, be the hyperplane $\{x \in E^n: x \cdot y = \eta(\omega, y)\}$. Since $z \cdot y \leq \eta(\omega, y)$ for all $z \in K(\omega)$ and $P(\omega, y) \cap K(\omega)$ is not empty, $P(\omega, y)$ is the <u>support hyperplane</u> of $K(\omega)$ with outward normal y. For each $y \neq 0$ the set $S(\omega, y) = P(\omega, y) \cap K(\omega)$ is called the <u>contact set</u> of $K(\omega)$ and its elements are called <u>contact points</u> of $K(\omega)$. It follows that $S(\omega, y)$ is not empty, $S(\omega, y) \subset \partial K(\omega)$, $S(\omega, \lambda y) = S(\omega, y)$ for $\lambda > 0$.

A function $s(\omega, y)$ defined on $\Omega \times E^n$ is a <u>contact function</u> of $K(\omega)$ if $s(\omega, y) \in S(\omega, y)$, $y \neq 0$, and $s(\omega, 0) \in K(\omega)$. Thus for $\omega \in \Omega$: $s(\omega, \cdot)$ is bounded; $s(\omega, y) = s(\omega, \lambda y)$, $\lambda > 0$; $\eta(\omega, y) = s(\omega, y) \cdot y$.

If for every $\omega \in \Omega$ and every $y \in E^n$ there is a method for determining a point $x(\omega, y) \in K(\omega)$ such that $x(\omega, y) \cdot y = \max_{z \in K(\omega)} z \cdot y = \eta(\omega, y)$, then it is said that a contact function of $K(\omega)$ is available. This availability is essential to the computing procedure GIP presented in the next section. Consider now the general problem:

GP <u>Let</u> $\Omega = [0, \hat{\omega}]$, $\hat{\omega} > 0$, <u>be a compact interval in</u> E^1 <u>and for</u> $\omega \in \Omega$ <u>let sets</u> $K(\omega) \subset E^n$ <u>be compact, convex, and continuous in</u> ω. <u>Assume that there exists</u> $\omega^* \in \Omega$ <u>such that</u> $0 \notin K(\omega)$, $0 \leq \omega < \omega^*$, <u>and</u> $0 \in K(\omega^*)$.
<u>Find</u>: ω^*.

Note that because $K(\omega)$ is continuous it follows that $0 \in \partial K(\omega^*)$. The relationship of GP to problems in optimal control is discussed in Chapter 5.

The computational problem posed by Fadden [Fl] differs from GP in that the sets $K(\omega)$ are required to be strictly convex. Furthermore, in addition to ω^* Fadden seeks to find the outward normal to a support hyperplane of $K(\omega^*)$ which contacts $K(\omega^*)$ at the point 0.

4.2 The General Iterative Procedure GIP

In this section the iterative procedure for computing the solution to GP is described.

First some preliminary definitions and results are presented. Let Ω and $K(\omega)$ be as specified in GP. For $\omega \in \Omega$ define functions $\tilde{z}(\omega)$ and $\rho(\omega) = |\tilde{z}(\omega)|$ such that $\tilde{z}(\omega) \in K(\omega)$ and $\rho(\omega) = |\tilde{z}(\omega)| = \min_{z \in K(\omega)} |z|$. The compactness of $K(\omega)$ and continuity of $|z|$ ensure the existence of these functions. Geometrically, $\tilde{z}(\omega)$ is the point in $K(\omega)$ closest to the origin and $\rho(\omega)$ is the distance of $\tilde{z}(\omega)$ from 0. It follows from Theorem 2.2.1 that: $\tilde{z}(\omega)$ is unique; $\rho(\omega) = 0$ if and only if $0 \in K(\omega)$; for $\rho(\omega) > 0$, $\tilde{z}(\omega) \in \partial K(\omega)$; for $\rho(\omega) > 0$, $z = \tilde{z}(\omega)$ if and only if $z \in P(\omega, -z) \cap K(\omega) = S(\omega, -z)$. Moreover, the following result holds.

THEOREM 4.2.1 <u>The functions</u> $\rho(\omega)$ <u>and</u> $\tilde{z}(\omega)$ <u>are continuous on</u> Ω.

<u>Proof</u>: Consider $\rho(\omega)$ first and let $\epsilon > 0$ be given. It is required to find $\delta_1 = \delta_1(\epsilon, \omega)$ such that $|\rho(\omega) - \rho(\omega + \bar{\omega})| < \epsilon$ whenever $|\bar{\omega}| < \delta_1$ where $\omega, \omega + \bar{\omega} \in \Omega$. Since $K(\omega)$ is continuous, there exists $\delta = \delta(\epsilon, \omega)$

such that $K(\omega) \subset K(\omega + \bar{\omega}) + N(0;\epsilon)$ and $K(\omega + \bar{\omega}) \subset K(\omega) + N(0;\epsilon)$ whenever $|\bar{\omega}| < \delta$. It will be shown that $\delta_1 = \delta$. The first inclusion and $\tilde{z}(\omega) \in K(\omega)$ imply that there exist $z_1 \in K(\omega + \bar{\omega})$ and $x_1 \in N(0;\epsilon)$ such that $\tilde{z}(\omega) = z_1 + x_1$. This yields $|z_1| = |\tilde{z}(\omega) - x_1| \leq |\tilde{z}(\omega)| + |x_1| < |\tilde{z}(\omega)| + \epsilon$. Similarly the second inclusion and $\tilde{z}(\omega + \bar{\omega}) \in K(\omega + \bar{\omega})$ imply that there exist $z_2 \in K(\omega)$ and $x_2 \in N(0;\epsilon)$ such that $\tilde{z}(\omega + \bar{\omega}) = z_2 + x_2$, $|z_2| < |\tilde{z}(\omega + \bar{\omega})| + \epsilon$. But $|\tilde{z}(\omega + \bar{\omega})| \leq |z_1|$ and $|\tilde{z}(\omega)| \leq |z_2|$. Therefore, $|\tilde{z}(\omega + \bar{\omega})| < |\tilde{z}(\omega)| + \epsilon$ and $|\tilde{z}(\omega)| < |\tilde{z}(\omega + \bar{\omega})| + \epsilon$, which yields the desired property: $|\rho(\omega) - \rho(\omega + \bar{\omega})| < \epsilon$ whenever $|\bar{\omega}| < \delta$.

Now consider the vector function $\tilde{z}(\omega)$. Let $\epsilon > 0$ be given and let $\delta = \delta(\epsilon, \omega)$ be chosen as in the preceding paragraph. The fact that $K(\omega)$ is bounded for all $\omega \in \Omega$ implies that there is an $\alpha > 0$ such that $\rho(\omega) < \alpha$, $\omega \in \Omega$. Since $|\tilde{z}(\omega) - \tilde{z}(\omega + \bar{\omega})| \leq \rho(\omega) + \rho(\omega + \bar{\omega}) < 2\alpha$, ϵ may be restricted to the interval $0 < \epsilon < 2\alpha$. Two cases will be considered, corresponding to $\rho(\omega) < \epsilon$ and $\rho(\omega) \geq \epsilon$.

Suppose first of all that $\rho(\omega) < \epsilon$. For $|\bar{\omega}| < \delta(\epsilon, \omega)$, $|\rho(\omega) - \rho(\omega + \bar{\omega})| < \epsilon$ which yields $\rho(\omega + \bar{\omega}) < \rho(\omega) + \epsilon < 2\epsilon$. Then $|\tilde{z}(\omega) - \tilde{z}(\omega + \bar{\omega})| \leq \rho(\omega) + \rho(\omega + \bar{\omega}) < 3\epsilon$.

Now consider $\rho(\omega) \geq \epsilon$ and $|\bar{\omega}| < \delta(\epsilon, \omega)$. Let $N_1 = N(0; \rho(\omega) + \epsilon)$, $Q_1 = Q(\tilde{z}(\omega); -\tilde{z}(\omega))$, $Q_2 = Q(\tilde{z}(\omega) - \epsilon\rho^{-1}(\omega)\tilde{z}(\omega); -\tilde{z}(\omega))$, and $Q_2^+ = Q^+(\tilde{z}(\omega) - \epsilon\rho^{-1}(\omega)\tilde{z}(\omega); -\tilde{z}(\omega))$. As illustrated in Figure 4.2.1 Q_1 is the support hyperplane of $K(\omega)$ at $\tilde{z}(\omega)$ with outward normal $-\tilde{z}(\omega)$. It can be shown that the parallel hyperplane Q_2 is the support hyperplane of $K(\omega) + \bar{N}(0;\epsilon)$ at $\tilde{z}(\omega) - \epsilon\rho^{-1}(\omega)\tilde{z}(\omega)$. Since $|\rho(\omega) - \rho(\omega + \bar{\omega})| < \epsilon$, $\tilde{z}(\omega + \bar{\omega}) \in N_1$. In addition, $\tilde{z}(\omega + \bar{\omega}) \in K(\omega) + N(0;\epsilon)$ implies $\tilde{z}(\omega + \bar{\omega}) \in Q_2^+$, the open half-space bounded by Q_2 with outward normal $-\tilde{z}(\omega)$. Thus if x is any point in $Q_2 \cap \partial N_1$ it follows that $|\tilde{z}(\omega) - \tilde{z}(\omega + \bar{\omega})| < |\tilde{z}(\omega) - x|$. Furthermore,

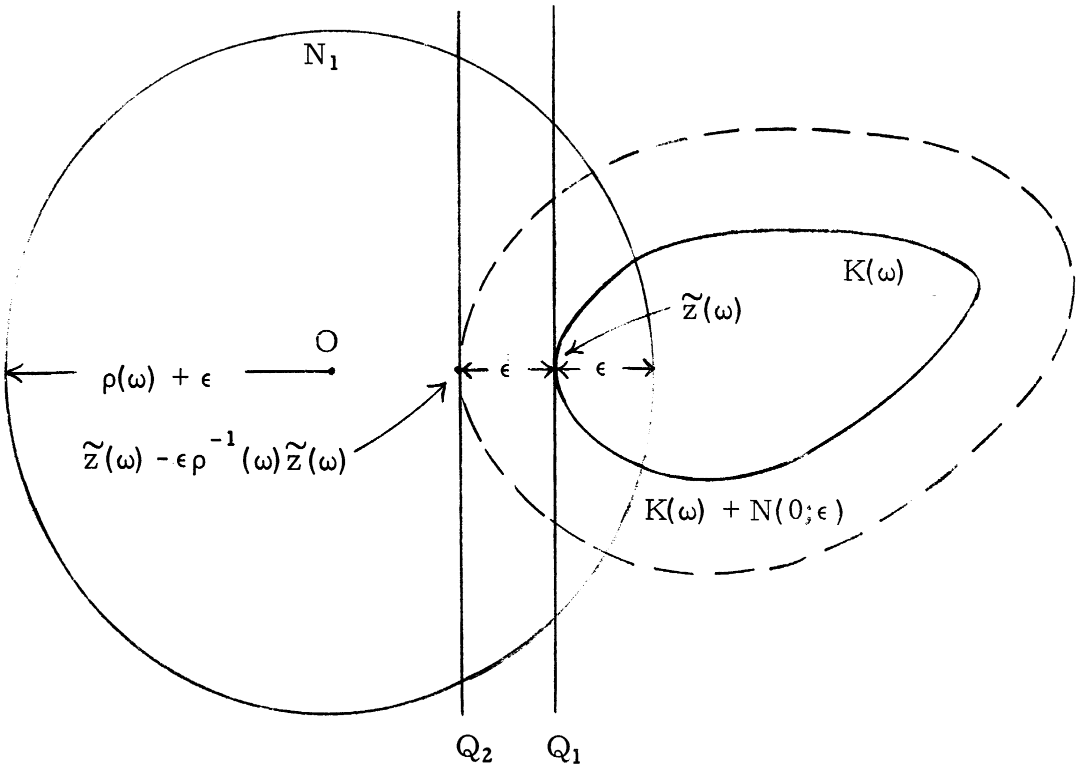


Figure 4. 2. 1 Notation for the case $\rho(\omega) \geq \epsilon$ in the proof of Theorem 4.2.1

$$\begin{aligned} |\tilde{z}(\omega) - x|^2 &= \epsilon^2 + [\rho(\omega) + \epsilon]^2 - [\rho(\omega) - \epsilon]^2 \\ &= \epsilon^2 + 4\rho(\omega)\epsilon \\ &= \epsilon\,[\epsilon + 4\rho(\omega)]. \end{aligned}$$

Since $\rho(\omega) < \alpha$ and $\epsilon < 2\alpha$, the quantity in the brackets is less than 6α. Thus $|\tilde{z}(\omega) - \tilde{z}(\omega + \bar{\omega})| < (6\alpha\epsilon)^{\frac{1}{2}}$.

Clearly the above results imply $|\tilde{z}(\omega) - \tilde{z}(\omega+\bar{\omega})| < \epsilon$ if $|\bar{\omega}| < \min\{\delta(\frac{\epsilon}{3}, \omega), \delta(\frac{\epsilon^2}{6\alpha}, \omega)\}$. Therefore $\tilde{z}(\omega)$ is continuous on Ω and the proof is complete.

Now let ω be an element of Ω and $s(\omega, \cdot)$ a specific contact function of $K(\omega)$. Consider

$$\gamma(\omega, z) = |z|^{-2} z \cdot s(\omega, -z), \quad |z| > 0, \quad z \cdot s(\omega, -z) > 0 \qquad (4.2.1)$$
$$= 0, \qquad z = 0 \text{ or } |z| > 0, \quad z \cdot s(\omega, -z) \leq 0.$$

Thus $\gamma(\omega, \cdot)$ is a function defined on $K(\omega)$. It follows from Theorem 2.3.1 that for $\omega \in \Omega$ and $z \in K(\omega)$: $0 \leq \gamma(\omega, z) \leq 1$; if $0 \in K(\omega)$, $\gamma(\omega, z) = 0$ for all $z \in K(\omega)$; if $0 \notin K(\omega)$, $\gamma(\omega, z) = 1$ if and only if $z = \tilde{z}(\omega)$; for fixed $\omega \in \Omega$, $\gamma(\omega, z)$ is continuous in z. The following results also hold.

THEOREM 4.2.2 Let Ω, $K(\omega)$, and ω^* be as specified in GP. Then for $0 \leq \omega < \omega^*$ and $z \in K(\omega)$: i) $|z|\gamma(\omega, z) \leq \rho(\omega)$; ii) $\eta(\omega^*, -y) \geq 0$, $y \in E^n$; iii) if $\gamma(\omega, z) > 0$, $\eta(\omega, -z) = -|z|^2 \gamma(\omega, z) < 0$.

Proof: Part i) follows from the definition of $\rho(\omega)$ and part i) of Theorem 3.3.1. Inequality (2.4.4) applied to the set $K(\omega^*)$ yields $s(\omega^*, -y) \cdot y \leq \tilde{z}(\omega^*) \cdot y$, $y \in E^n$. Since $\tilde{z}(\omega^*) = 0$, $\eta(\omega^*, -y) = -y \cdot s(\omega^*, -y) \geq 0$ which proves ii). Consider iii). The fact that $0 \notin K(\omega)$ for $0 \leq \omega < \omega^*$ implies $|z| > 0$, all $z \in K(\omega)$. This, $\gamma(\omega, z) > 0$, and (4.2.1) yield $z \cdot s(\omega, -z) > 0$. So iii) follows from $\eta(\omega, -z) = -z \cdot s(\omega, -z)$ and (4.2.1).

Now GIP can be stated.

The General Iterative Procedure GIP Let Ω, $K(\omega)$, and ω^* be as described in GP. For $\omega \in \Omega$ let $\eta(\omega, \cdot)$ be the support function of $K(\omega)$, let $s(\omega, \cdot)$ be an arbitrary contact function of $K(\omega)$, and define $\gamma(\omega, \cdot)$ by (4.2.1). Set $\omega_0 = 0$ and $i = 0$. The following two steps constitute one iteration, called iteration i, of GIP.

Step 1 Consider ω_i fixed, $0 \leq \omega_i \leq \omega^*$, and the corresponding set $K(\omega_i)$. Apply IIP or BIP to the minimization of $|z|$, $z \in K(\omega_i)$. If $\omega_i < \omega^*$, IIP or BIP may be continued (Theorems 3.4.1 and 2.4.1) until a point $z_i \in K(\omega_i)$ is obtained such that $\gamma(\omega_i, z_i) \geq \theta_i$, where $0 < \theta \leq \theta_i < 1$ and θ, θ_i are preselected numbers. When this happens proceed to Step 2. If $\omega_i = \omega^*$, IIP or BIP will not produce a $z \in K(\omega_i)$ satisfying $\gamma(\omega_i, z) \geq \theta_i$ but will generate (Theorems 3.4.1 and 2.4.1) a sequence of points in $K(\omega_i) = K(\omega^*)$ which converges to 0.

Step 2 Consider z_i fixed and look at $\eta(\omega, -z_i)$ for $\omega \in [\omega_i, \omega^*]$. Since $\gamma(\omega_i, z_i) \geq \theta$, $\eta(\omega_i, -z_i) < 0$ (Theorem 4.2.2). This, $\eta(\omega^*, -z_i) \geq 0$, and the continuity of $\eta(\omega, z)$ (Theorem 4.1.1) imply there exists $\bar{\omega} \in (\omega_i, \omega^*]$ such that $\eta(\bar{\omega}, -z_i) = 0$. Let ω increase from ω_i and let ω_{i+1} be the first ω for which $\eta(\omega, -z_i) = 0$. Then increase i by one and return to Step 1.

It should be noted that in Step 1 with $i > 0$ the point $s(\omega_i, -z_{i-1})$ lying on the hyperplane $Q(0; -z_{i-1})$ can be used to initiate IIP or BIP. For $i = 0$ the point $s(\omega_i, x)$ for any x can be used.

If the condition $\omega_i = \omega^*$ in the latter part of Step 1 occurs, then the solution to GP has been obtained in a finite number of iterations of GIP. If this condition does not occur, then GIP generates infinite sequences $\{\omega_i\}$ and $\{z_i\}$ where $\omega_i \in [0, \omega^*)$ and $z_i \in K(\omega_i)$. To show convergence of ω_i to ω^* and z_i to 0 it is necessary to make an additional assumption on $\eta(\omega, y)$, namely, that the difference quotient $\dfrac{\eta(\omega_a, y) - \eta(\omega_b, y)}{\omega_a - \omega_b}$ be bounded from above for all $\omega_a \omega_b \in \Omega$, $\omega_a \neq \omega_b$, bounded $y \in E^n$. This is treated in the next section.

4.3 Convergence Theorem for GIP

THEOREM 4.3.1 Consider GIP and assume the condition $\omega_i = \omega^*$ in the latter part of Step 1 does not occur for $i < \infty$. Furthermore, assume that $\dfrac{\eta(\omega_a, y) - \eta(\omega_b, y)}{\omega_a - \omega_b}$ is bounded from above for all $\omega_a, \omega_b \in \Omega$, $\omega_a \neq \omega_b$, bounded $y \in E^n$. Then GIP generates sequences $\{\omega_i\}$ and $\{z_i\}$ which for $i \geq 0$ and $i \to \infty$ satisfy: i) $\omega_i \in [0, \omega^*)$ and $z_i \in K(\omega_i)$; ii) $\{\omega_i\}$ is strictly increasing ($\omega_i < \omega_{i+1}$) and $\omega_i \to \omega^*$; iii) $z_i \to 0$.

Proof: The first half of i) follows from $\omega_0 \in [0, \omega^*)$ and the definition of ω_{i+1} in Step 2. Since all points z generated by IIP or BIP applied to the set $K(\omega_i)$ are in $K(\omega_i)$, $z_i \in K(\omega_i)$.

Consider ii). The property $\omega_i < \omega_{i+1}$ follows by the remarks in Step 2. Thus $\{\omega_i\}$ is a strictly increasing sequence on a compact interval $[0, \omega^*]$ and must have a limit, say $\tilde{\omega}$. To complete the proof of ii) it suffices to show that $\omega^* - \epsilon \le \tilde{\omega} \le \omega^*$, where $\epsilon > 0$ is chosen arbitrarily.

Assume the contrary, i.e., $\tilde{\omega} < \omega^* - \epsilon$. Let

$$\bar{\nu} = \underset{\substack{\omega_a, \omega_b \in \Omega, \omega_a \neq \omega_b \\ -y \in \bigcup_{\omega \in \Omega} K(\omega)}}{\text{supremum}} \frac{\eta(\omega_a, y) - \eta(\omega_b, y)}{\omega_a - \omega_b},$$

which exists by hypothesis. Clearly $\bar{\nu} > 0$. At the start of Step 2 $\omega_i < \omega^*$, $|z_i| > 0$, and $\gamma(\omega_i, z_i) \ge \theta$. By part iii) of Theorem 4.2.2, $\eta(\omega_i, -z_i) = -|z_i|^2 \gamma(\omega_i, z_i) < 0$. But $\eta(\omega_{i+1}, -z_i) = 0$. Hence,

$$\omega_{i+1} - \omega_i \ge \bar{\nu}^{-1} |z_i|^2 \gamma(\omega_i, z_i) \ge \bar{\nu}^{-1} |z_i|^2 \theta > 0. \tag{4.3.1}$$

Now consider $\omega \in [0, \omega^* - \epsilon]$ and the function $\tilde{z}(\omega)$. Since $\tilde{z}(\omega)$ is continuous on $[0, \omega^* - \epsilon]$ (Theorem 4.2.1), it takes on its minimum, say $\tilde{z}_{min}$, on this interval. Then $|\tilde{z}_{min}| = \min_{0 \le \omega \le \omega^* - \epsilon} [\min_{z \in K(\omega)} |z|] > 0$. Consequently $|z_i| \ge |\tilde{z}_{min}|$ for all $i \ge 0$ which by (4.3.1) implies $\omega_{i+1} - \omega_i \ge \bar{\nu}^{-1} |\tilde{z}_{min}|^2 \theta = \text{constant} > 0$ for all $i \ge 0$. This contradicts the fact that $\{\omega_i\}$ has a limit on $[0, \omega^* - \epsilon]$ and thus ii) is true.

Now consider iii). From $\omega_i \to \omega^*$, the continuity of $\tilde{z}(\omega)$, and $\tilde{z}(\omega^*) = 0$, it follows that $\tilde{z}(\omega_i) \to 0$. Part i) of Theorem 4.2.2 and $\gamma(\omega_i, z_i) \ge \theta > 0$ yield $0 \le |z_i|\theta \le |z_i|\gamma(\omega_i, z_i) \le |\tilde{z}(\omega_i)|$. Since the right side of this inequality converges to 0, $|z_i|\theta \to 0$ and thus $|z_i| \to 0$. That is, $z_i \to 0$, which completes the proof.

Note that it is not possible to say that $|z_{i+1}| \leq |z_i|$ (see Figure 4. 3. 1).

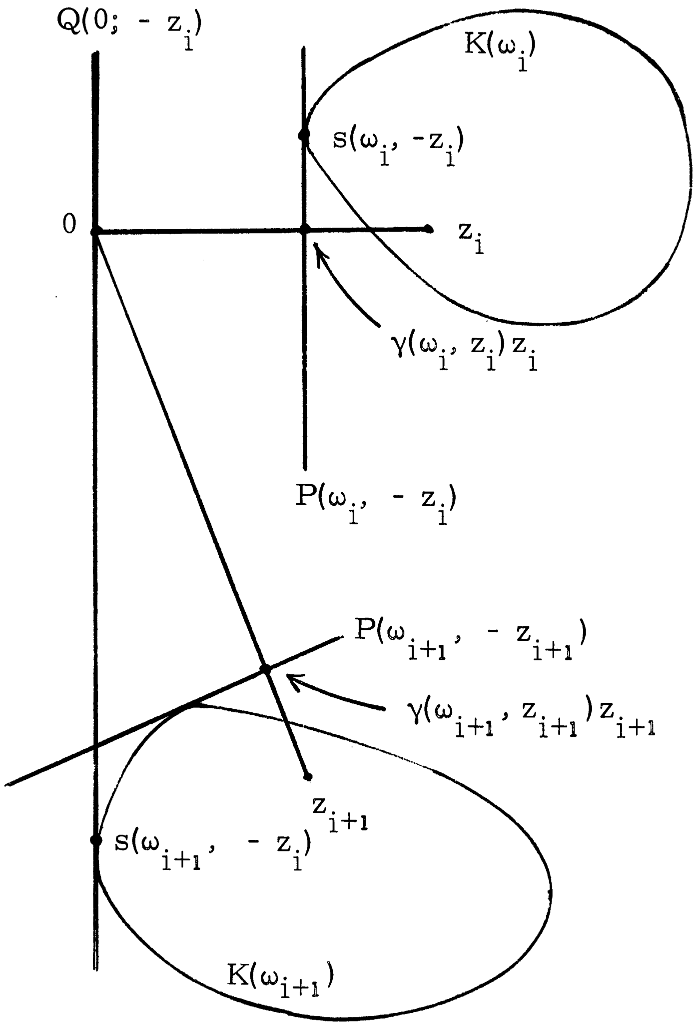


Figure 4. 3. 1 A configuration for which $|z_{i+1}| > |z_i|$ in GIP.

Also observe that if on an iteration of IIP or BIP within Step 1 of GIP a point $z \in K(\omega_i)$ is obtained for which $|z| = 0$, then $\omega_i = \omega^*$. However, if $|z| < \epsilon$ (arbitrary $\epsilon > 0$) is used as a stopping condition for GIP, then ω_i may or may not be close to ω^*. In certain applications where ω represents fuel, cost, effort, time, etc. (see Chapter 5) it may be satisfactory to obtain $\tilde{\omega}$, $0 \leq \tilde{\omega} \leq \omega^*$, provided that a point $z \in K(\tilde{\omega})$ sufficiently near to the origin is found. In such cases $|z| < \epsilon$ provides a stopping condition for GIP. Consider the following:

COROLLARY 4.3.1 Let the hypotheses of Theorem 4.3.1 be satisfied, consider $\{\omega_i\}$ and $\{z_i\}$ generated by GIP, and let $\epsilon > 0$ be given. Then: i) if $\omega^* - \omega_i < \epsilon$, $|z_i| < (\bar{\nu}\theta^{-1}\epsilon)^{\frac{1}{2}}$; ii) if $|z_i| < \epsilon$ and for $\omega_a, \omega_b \in [0, \omega^*]$, $\omega_a \neq \omega_b$, $\dfrac{\rho(\omega_a) - \rho(\omega_b)}{\omega_a - \omega_b} \leq -\sigma < 0$, it follows that $\omega^* - \omega_i < \sigma^{-1}\epsilon$.

Proof: Inequality (4.3.1), $\omega_i < \omega_{i+1} < \omega^*$, and $\omega^* - \omega_i < \epsilon$ imply $\bar{\nu}^{-1}|z_i|^2\theta < \epsilon$ which proves i). Consider ii). For $\omega_a = \omega^*$, $\omega_b = \omega_i$ it follows: $\dfrac{\rho(\omega^*) - \rho(\omega_i)}{\omega^* - \omega_i} \leq -\sigma$. Since $\rho(\omega^*) = 0$ and $\omega_i < \omega^*$, $-\rho(\omega_i) \leq -\sigma(\omega^* - \omega_i)$ and $\omega^* - \omega_i \leq \sigma^{-1}\rho(\omega_i)$. But $\rho(\omega_i) \leq |z_i| < \epsilon$ so $\omega^* - \omega_i < \sigma^{-1}\epsilon$, completing the proof.

Part ii) of the corollary has importance if the scalar $\sigma > 0$ exists and can be calculated. For in that case whenever the stopping condition $|z| < \epsilon$ is satisfied for a point $z \in K(\omega_i)$ generated by IIP or BIP within Step 1 of GIP, ii) provides a measure of the error in ω_i. Moreover, if $|z| < \sigma\epsilon$ is used as the stopping condition for GIP, then $\omega^* - \omega_i < \epsilon$.

CHAPTER 5

APPLICATION OF BIP, IIP, AND GIP TO OPTIMAL CONTROL PROBLEMS

5.1 Some Optimal Control Problems Solvable by GIP

In this section a class of optimal control problems is formulated using some ideas of Fadden [Fl]. Many problems of this class, and other optimal control problems as well, can be solved by GIP. Some can be solved by BIP or IIP alone. Six specific problems which illustrate a variety of optimization objectives are stated. The application of the iterative procedures to these six problems is discussed later in the chapter.

Let U be a compact set in E^r and let $\Theta = [0, \hat{t}]$, $\hat{t} > 0$, be a compact interval in E^1. A measurable function $u(\cdot)$ defined on Θ whose range is in U is said to be an admissible control.

Consider dynamical systems of the form

$$\dot{x}(t) = A(t)\,x(t) + f(u(t), t), \qquad x(0) \text{ specified,} \tag{5.1.1}$$

where x is the m-dimensional state vector, $\dot{x}$ is its time derivative, $x(0)$ is the initial state, $u(\cdot)$ is an r-dimensional vector control function defined on Θ, $A(\cdot)$ is an $m \times m$ matrix function defined and continuous on Θ, $f(\cdot, \cdot)$ is an m-dimensional vector function defined and continuous on $U \times \Theta$. For every admissible control $u(\cdot)$ there is an absolutely continuous solution function $x_{u(\cdot)}(t) = x_u(t)$, which satisfies (5.1.1) almost everywhere on Θ. An admissible $u(\cdot)$ is said to generate the states $x_u(t)$.

Let there be given target sets $W(t) \subset E^m$ which are compact, convex, and continuous on Θ. An admissible control $u(\cdot)$ is said to transfer x from $x(0)$ to $W(t)$ in time $t \in \Theta$ if $x_u(t) \in W(t)$.

Let there also be given a cost functional of the form

$$J_t(u) = \int_0^t [a(\sigma) \cdot x_u(\sigma) + f^0(u(\sigma), \sigma)]d\sigma + h^0(x_u(t)), \qquad (5.1.2)$$

where $a(\cdot)$ is a continuous function from Θ to E^m, $f^0(\cdot,\cdot)$ is a continuous function from $U \times \Theta$ to E^1, and $h^0(\cdot)$ is a convex (continuous) function from E^m to E^1. For every admissible control $u(\cdot)$ $J_t(u)$ yields a particular cost function, which if evaluated at $t \in \Theta$ results in a number called the cost at time t for this admissible control.

Consider now the following class of optimal control problems:

Fixed Terminal Time Problems

Let $T > 0$ be a fixed point in Θ. Find an admissible control $u^*(\cdot)$, if one exists, such that: a) $u^*(\cdot)$ transfers x from $x(0)$ to $W(T)$ in time T; b) the cost at time T for $u^*(\cdot)$ is less than or equal to the cost at time T for any admissible control $u(\cdot)$ which transfers x from $x(0)$ to $W(T)$ in time T. (For these problems $J_t(u) = J_T(u) \equiv J(u)$.)

Free Terminal Time Problems

Find an admissible control $u^*(\cdot)$ and an optimal time $t^* \in \Theta$, if these exist, such that: a) $u^*(\cdot)$ transfers x from $x(0)$ to $W(t^*)$ in time t^*; b) the cost at time t^* for $u^*(\cdot)$ is less than or equal to the cost at time $\tilde{t}$ for any admissible control $\tilde{u}(\cdot)$ and time $\tilde{t} \in \Theta$ for which $x_{\tilde{u}}(\tilde{t}) \in W(\tilde{t})$.

Any admissible control $u^*(\cdot)$ which satisfies conditions a) and b) in either a fixed or free terminal time problem is said to be an optimal control.

The above formulation is also appropriate if there are constraints on the control function of the form

$$\int_0^t \phi^i(u(\sigma), \sigma)d\sigma \in M^i, \quad i = 1, 2, \ldots, \ell, \qquad (5.1.3)$$

where the $\phi^i(\cdot,\cdot)$ are continuous functions from $U \times \Theta$ to E^1 and M^i are closed intervals in E^1. By introducing ℓ additional differential equations

$$\dot{x}^{(m+i)} = \phi^i(u(t), t), \quad \dot{x}^{(m+i)}(0) = 0, \quad i = 1, 2, \ldots, \ell, \qquad (5.1.4)$$

and letting $(x(t), x^{(m+1)}(t), \ldots, x^{(m+\ell)}(t)) \in E^{m+\ell}$ be the new state vector, these constraints can be incorporated into a system description of the form of equation (5.1.1). The target sets in this case are $W(t) \times M^1 \times \cdots \times M^\ell$. Since there is no loss of generality, it is assumed in the sequel that no constraints of this type are present.

Many optimal control problems of the class described above can be solved by GIP. Furthermore, it is often possible to use GIP on problems of an extended class which includes other cost functionals (e.g., equation (5.1.9)) and/or target sets which are closed but not bounded (e.g., half-spaces or linear manifolds of dimension ≥ 1 which may or may not vary with t). Some of these problems can be solved by BIP or IIP alone.

The following six example problems will be used later in the chapter to illustrate the application of the iterative procedures. The first five have a fixed terminal time T; the sixth has a free terminal time.

<u>Problem 1 (Minimum-Error Regulator)</u>

$$J(u) = |x_u(T)| \qquad (5.1.5)$$

For this problem, first posed by Ho [H3], the target set W(T) is equal to the whole space E^m.

<u>Problem 2 (Generalized Minimum-Error Regulator)</u>

$$J(u) = x_u(T) \cdot Gx_u(T) + g \cdot x_u(T), \qquad (5.1.6)$$

where G is a symmetric non-negative definite $m \times m$ matrix and g is an m-vector in the range of G. This problem, treated by Gilbert [Gl], also has the target set $W(T) = E^m$.

Problem 3 (Convex Function Minimum-Error Regulator)

$$J(u) = h^0(x_u(T)), \tag{5.1.7}$$

where $h^0(\cdot)$ is a convex (continuous) function from E^m to E^1 such that $\{x : h^0(x) \le \text{constant}\}$ is compact, and $W(T) = E^m$.

Problem 4 (Minimum-Fuel Terminal Control)

$$J(u) = \int_0^T [a(t) \cdot x_u(t) + f^0(u(t), t)]\,dt\,, \tag{5.1.8}$$

where $a(\cdot)$ is a continuous function from $[0, T]$ to E^m and $f^0(\cdot,\cdot)$ is a continuous function from $U \times [0, T]$ to E^1.

Problem 5 (Minimum Effort)

$$J(u) = \sup_{\substack{0 \le t \le T \\ 1 \le i \le r}} |u^i(t)|\,. \tag{5.1.9}$$

For this problem, which was introduced by Neustadt [N3], the target set $W(T)$ is a single point $w(T)$ and $f(u, t) = B(t)u$ where $B(\cdot)$ is an $m \times r$ matrix function continuous on $[0, T]$.

Problem 6 (Minimum Time)

$$J_t(u) = t, \tag{5.1.10}$$

which corresponds to (5.1.2) with $a(\cdot) = 0$, $f^0(\cdot,\cdot) = 1$ and $h^0(\cdot) = 0$. There are no further restrictions on $W(t)$ for this problem.

5.2 The Reachable Set; Determination of a Contact Function

The problem of computing an optimal control can be approached by considering the set of all possible solutions of (5.1.1) using admissible controls. This set and some of its important properties are discussed in this section. In addition, two closely-related sets are introduced.

Consider the system (5.1.1) and let $t \in \Theta$. The set

$$R(t) = \{x : x = x_u(t),\ u(\cdot) \text{ admissible}\} \tag{5.2.1}$$

is called the reachable set. It is the set of all possible states in E^m which can be reached from x(0) at time t using an admissible control.

Since the solution to (5.1.1) can be written as

$$x_u(t) = \Phi(t)\left[x(0) + \int_0^t \Phi^{-1}(\sigma)\, f(u(\sigma), \sigma)\, d\sigma\right] \tag{5.2.2}$$

where $\Phi(t)$ is the $m \times m$ matrix solution of $\dot{\Phi} = A(t)\Phi$, $\Phi(0) = I$, it is clear that

$$R(t) = \left\{\Phi(t)\left[x(0) + \int_0^t \Phi^{-1}(\sigma)\, f(u(\sigma), \sigma)\, d\sigma\right] : u(\cdot) \text{ admissible}\right\}. \tag{5.2.3}$$

Neustadt has shown [N2] that for each $t \in \Theta$ the set R(t) is: (1) compact; (2) convex. From (5.2.3) and the continuity of $f(\cdot, \cdot)$ and $\Phi(\cdot)$ it is easy to prove that: (3) R(t) is continuous on Θ. Furthermore, for fixed $t \in \Theta$ it will be shown below that: (4) a contact function of the set R(t) is available. The properties (1), (2), (3), (4) are the essential features which permit application of the computing procedures of Chapters 2, 3, and 4 to optimal control problems.

It should be observed that for all but the most elementary systems (5.1.1) the complexity of (5.2.3) prohibits explicit calculation of the boundaries of R(t). Hence, the most satisfactory computing algorithms are those based solely on properties (1), (2), (3), (4).

Let $\tau \in \Theta$ be given and consider the determination of a contact function $s_R(\tau, y)$, $y \in E^m$, of $R(\tau)$. For y a specified m-vector it follows from (5.2.2) that

$$y \cdot x_u(\tau) = y \cdot \Phi(\tau)\, x(0) + \int_0^\tau ([\Phi(\tau)\Phi^{-1}(\sigma)]' y) \cdot f(u(\sigma),\sigma)\, d\sigma, \qquad (5.2.4)$$

where the prime denotes matrix transpose. But the m-vector $\psi(t, y) = [\Phi(\tau)\Phi^{-1}(t)]' y$, defined on $[0, \tau] \times E^m$, is the solution of the adjoint differential equation

$$\dot{\psi}(t) = -A'(t)\, \psi(t), \quad \psi(\tau) = y. \qquad (5.2.5)$$

Hence,

$$\max_{x \in R(\tau)} y \cdot x = \psi(0, y) \cdot x(0) + \int_0^\tau \max_{v \in U} [\psi(\sigma, y) \cdot f(v, \sigma)]\, d\sigma. \qquad (5.2.6)$$

It can be shown that there exists an admissible control $u(t, y)$ defined on $[0, \tau] \times E^m$ such that for almost all $t \in [0, \tau]$:

$$\psi(t, y) \cdot f(u(t, y), t) = \max_{v \in U} \psi(t, y) \cdot f(v, t). \qquad (5.2.7)$$

Then from (5.2.6) it is clear that $y \cdot x_{u(t,y)}(\tau) \geq y \cdot x_{u(t)}(\tau)$ for every admissible control $u(t)$. It follows that a contact function of $R(\tau)$ is

$$s_R(\tau, y) = x_{u(t,y)}(\tau). \qquad (5.2.8)$$

This result agrees with the well-known fact that boundary points of the reachable set must "satisfy" the Pontryagin maximum principle.

In most practical problems it is not difficult to obtain a function $u(t, y)$ which satisfies (5.2.7). Consider, for example, the case where $f(u, t) = B(t)u$, $B(t)$ is an $m \times r$ matrix function continuous on $[0, \tau]$, and U is the unit hypercube $\{u : |u^i| \leq 1,\ i = 1, 2, \ldots, r\}$. Notice that (5.2.7) may not uniquely define $u(t, y)$ almost everywhere in $[0, \tau]$. For in the present example, (5.2.7) yields $u(t, y) = \operatorname{sgn} B'(t)\, \psi(t, y)$ where the ith component of sgn v $(i = 1, 2, \ldots, r)$ is 1 for $v^i > 0$, is -1 for $v^i < 0$, and is arbitrary if $v^i = 0$. This is of no concern, however, since even if a component of $B'(t)\, \psi(t, y)$ is identically zero on $[0, \tau]$ (i.e., LaSalle's normality condition [L1] is not satisfied) different choices for $u(t, y)$ will at most lead only to different contact functions of $R(\tau)$. Previous computational procedures [B1, E1, F1, F2, N1, N3, N4] have required assumptions which correspond to a unique determination of $u(t, y)$ by (5.2.7). Such "unique maximum" assumptions imply strict convexity of $R(\tau)$.

Computer evaluation of $s_R(\tau, y)$ for a specified $y \in E^m$ entails three steps:

i) Evaluation of $\psi(t, y)$ by solving (5.2.5) backwards from $t = \tau$ to $t = 0$;

ii) Determination of $u(t, y)$ from $\psi(t, y)$ by (5.2.7);

iii) Solution of (5.1.1) with $u(t) = u(t, y)$ from $t = 0$ to $t = \tau$.

As observed by Fancher [F3], storage of the function $u(t, y)$ can be avoided by first obtaining $\psi(0, y)$ from the backward integration of (5.2.5) and then integrating (5.2.5) and (5.1.1) together from $t = 0$ to $t = \tau$ while obtaining $u(t, y)$ from (5.2.7). Thus each evaluation of a contact function involves the sequential solution of two differential equations. This situation can be handled effectively by a digital or hybrid computer [G2].

It is convenient to introduce here two sets $C(t)$ and $\underline{R}(t)$ which are closely related to $R(t)$ and have application to a number of optimal control problems.

Let $t \in \Theta$ and define

$$C(t) = \left\{ \int_0^t \Phi^{-1}(\sigma)\, f(u(\sigma),\sigma)\, d\sigma \ : \ u(\cdot) \text{ admissible} \right\}. \qquad (5.2.9)$$

Note that $C(t) \subset E^m$ is related to $R(t)$ by the continuous, non-singular affine transformation

$$C(t) = \Phi^{-1}(t)\, R(t) - x(0). \qquad (5.2.10)$$

Consequently, $C(t)$ is compact, convex, and continuous in t for all $t \in \Theta$. Furthermore, a contact function $s_C(t, y)$, $y \in E^m$, of $C(t)$ is available, as is shown in the following paragraph.

Let $\tau \in \Theta$ and $y \in E^m$ be specified. If $c_u(\tau)$ denotes the point in $C(\tau)$ generated by the admissible control $u(t)$, then:

$$y \cdot c_u(\tau) = \int_0^\tau ([\Phi^{-1}(\sigma)]' y) \cdot f(u(\sigma), \sigma)\, d\sigma \qquad (5.2.11)$$

and

$$\max_{c \in C(\tau)} y \cdot c = \int_0^\tau \max_{v \in U} ([\Phi^{-1}(\sigma)]' y) \cdot f(v, \sigma)\, d\sigma \,. \qquad (5.2.12)$$

It can be shown that there exists an admissible control $u(t, y)$ defined on $[0, \tau] \times E^m$ such that for almost all $t \in [0, \tau]$:

$$([\Phi^{-1}(t)]' y) \cdot f(u(t, y), t) = \max_{v \in U} ([\Phi^{-1}(t)]' y) \cdot f(v, t)\,. \qquad (5.2.13)$$

Then if

$$s_C(\tau, y) = \int_0^\tau \Phi^{-1}(\sigma)\, f(u(\sigma, y), \sigma)\, d\sigma \,, \qquad (5.2.14)$$

it is clear that $y \cdot s_C(\tau, y) \geq y \cdot c_u(\tau)$ for every admissible control $u(t)$. Thus $s_C(\tau, y)$ is a contact function of $C(\tau)$. An alternate formula for evaluating $s_C(\tau, y)$ is

$$s_C(\tau, y) = \Phi^{-1}(\tau)\, x_{u(t,y)}(\tau) - x(0)\,, \qquad (5.2.15)$$

where $u(t, y)$ is obtained from (5.2.13). This follows directly from (5.2.10) and (5.2.14).

In certain optimal control problems such as Problem 4 of Section 5.1 it is useful to introduce a set $\underline{R}(t)$. For these problems there is a cost function $x_u^0(t) \in E^1$ satisfying almost everywhere on Θ the differential equation

$$\dot{x}^0(t) = a(t) \cdot x_u(t) + f^0(U(t), t), \quad x^0(0) = 0, \tag{5.2.16}$$

where $a(\cdot)$ and $f^0(\cdot, \cdot)$ are as in (5.1.2). Then if $\underline{x} = (x^0, x)$ denotes a point in E^{m+1}, (5.1.1) and (5.2.16) are equivalent to the equation

$$\dot{\underline{x}}(t) = \underline{A}(t)\,\underline{x}(t) + \underline{f}(u(t), t), \quad \underline{x}(0) = (0, x(0)), \tag{5.2.17}$$

where $\underline{f} = (f^0, f)$ and $\underline{A}(t)$ is given by

$$\underline{A}(t) = \left[\begin{array}{c|c} 0 & a^1(t) \quad a^2(t) \cdots a^m(t) \\ \hline 0 & \\ \vdots & A(t) \\ 0 & \end{array}\right]$$

Let $\underline{x}_u(t)$ be an absolutely continuous solution of (5.2.17) corresponding to the admissible control $u(\cdot)$. Then for $t \in \Theta$ define

$$\underline{R}(t) = \{\underline{x} : \underline{x} = \underline{x}_u(t), \ u(\cdot) \text{ admissible}\}. \tag{5.2.18}$$

If $\underline{\Phi}(t)$ is the $(m+1) \times (m+1)$ matrix solution of $\dot{\underline{\Phi}} = \underline{A}(t)\,\underline{\Phi}$, $\underline{\Phi}(0) = I$, $\underline{y}$ is a specified $(m+1)$-vector, and $\underline{\psi}(t, \underline{y})$, defined on $[0, \tau] \times E^{m+1}$, is the solution of

$$\dot{\underline{\psi}}(t) = -\underline{A}'(t)\,\underline{\psi}(t), \quad \underline{\psi}(\tau) = \underline{y}, \tag{5.2.19}$$

then equations (5.2.2) through (5.2.8) may be rewritten with sub-bars on each x, Φ, f, R, y, A, and ψ. Thus $\underline{R}(t)$ is compact, convex, and continuous on Θ. Moreover, for $\tau \in \Theta$, a contact function of $\underline{R}(\tau)$ is

$$s_{\underline{R}}(\tau, \underline{y}) = \underline{x}_{u(t,\underline{y})}(\tau), \tag{5.2.20}$$

where $u(t, \underline{y})$ satisfies for almost all $t \in [0, \tau]$:

$$\underline{\psi}(t, \underline{y}) \cdot f(u(t, \underline{y}), t) = \max_{v \in U} \underline{\psi}(t, \underline{y}) \cdot \underline{f}(v, t). \tag{5.2.21}$$

5.3 Solution of Problems 1 and 2 by BIP or IIP Alone

Problems 1 and 2 of Section 5.1 can be solved by GIP or by BIP or IIP alone. The method of solution by BIP or IIP alone is treated in this section.

In Problem 1 let $K = R(T)$, where $T > 0$ is the fixed terminal time. The optimization objective is: find a point $z^* \in K$ and an admissible control $u^*(\cdot)$ such that $|z^*| = \min_{z \in K} |z|$ and $z^* = x_{u^*}(T)$. Since $R(T)$ is compact, convex, and has an available contact function (equation (5.2.8)), this is clearly a variant of BP and BIP or IIP may be used to obtain a point $\tilde{z} \in K$ such that $|\tilde{z} - z^*| < \epsilon$ (arbitrary $\epsilon > 0$). There is no difficulty in initiating the iterative procedures: choose $z_0 = x_{u_0}(T) \in K$, where $u_0(\cdot)$ is any admissible control. (In IIP with Selection Rule D, an arbitrary admissible control say $u_{-1}(\cdot)$, can be used to generate $z_{-1} \in K$.) Furthermore, for all $k \geq 0$ the contact point $s(-z_k) = s_R(T, -z_k)$ is generated by an admissible control $u(t, -z_k)$ satisfying (5.2.7).

There remains the issue of finding admissible controls $u_k(\cdot)$ and $\tilde{u}(\cdot)$ which will generate z_k and $\tilde{z}$. For BIP or IIP (with Selection Rules A, B, C, or D) z_k, $k > 0$, has a representation: $z_k = \sum_{i=0}^{k-1} \sigma_i s(-z_i) + \sigma_{-1} z_0$ where $\sigma_i \geq 0$ $(i = -1, 0, \ldots, k-1)$ and $\sum_{i=-1}^{k-1} \sigma_i = 1$. Suppose that for almost all $t \in [0, T]$ the sets $f(U, t) = \{f(v, t) : v \in U\}$ are convex. Then for $k > 0$ there exists an admissible control $u_k(\cdot)$ such that almost everywhere in $[0, T]$ $f(u_k(t), t) = \sum_{i=0}^{k-1} \sigma_i f(u(t, -z_i), t) + \sigma_{-1} f(u_0(t), t)$, which implies by (5.1.1) that $z_k = x_{u_k(t)}(T)$. For example, if $f(u, t) = B(t)u + b(t)$ and U is convex, then $u_k(t) = \sum_{i=0}^{k-1} \sigma_i u(t, -z_i) + \sigma_{-1} u_0(t)$. If the sets $f(U, t)$ are not convex for almost all $t \in [0, T]$ and additional approximation process, the construction of a chattering control, is necessary [B3]. This chattering process yields an admissible control $\bar{u}_k(\cdot)$ which generates a point arbitrarily close to z_k. As a brief illustration, for $f(u, t) = B(t)u + b(t)$ and U not convex, the process is as follows.

Consider $\hat{u}(t) = \sum_{i=-1}^{k-1} \sigma_i v_i(t)$ where the σ_i are as above and the admissible functions $v_i(t)$ are: $v_{-1}(t) = u_0(t)$, $v_i(t) = u(t, -z_i)$, $i = 0, 1, \ldots, k-1$. Suppose $\hat{u}(t)$ does not have its range in U. Since $\sum_{i=-1}^{k-1} \sigma_i = 1$ and $\sigma_i \geq 0$ $(i = -1, 0, \ldots, k-1)$, the intervals

$$\Gamma_{rs} = \left((r-1)\tau + \sum_{i=-1}^{s} \sigma_i \tau - \sigma_s \tau,\ (r-1)\tau + \sum_{i=-1}^{s} \sigma_i \tau\right],$$

$\tau = q^{-1} T$, $r = 1, 2, \ldots, q$, $s = -1, 0, \ldots, k-1$, are mutually disjoint and cover $(0, T]$. Thus $\bar{u}_k(t) = v_s(t)$, $t \in \Gamma_{rs}$, $r = 1, 2, \ldots, q$, $s = -1, 0, \ldots, k-1$, is an admissible control defined on $(0, T]$ which "chatters" among $v_{-1}(t), v_0(t), \ldots, v_{k-1}(t)$, the fractional time spent on $v_i(t)$ being σ_i. It is not difficult to show that there exists a q_ϵ (dependent on ϵ, $A(t)$, $B(t)$, $b(t)$, T, and U) such that $|x_{\hat{u}}(T) - x_{\bar{u}_k}(T)| < \epsilon$ for $q > q_\epsilon$. Thus a sufficiently fine division of $(0, T]$ implies $|x_{\hat{u}}(T) - x_{\bar{u}_k}(T)| < \hat{\epsilon}$ (arbitrary $\hat{\epsilon} > 0$). It is clear that chattering can be omitted on any subinterval of $(0, T]$ where $\hat{u}(t)$ has range in U.

By Theorems 2.4.1 and 3.4.1 $z_k \to z^*$ for BIP and IIP. Thus for k sufficiently large z_k will satisfy specified stopping criteria. The approximate optimal state is $\tilde{z} = z_k$ and it is generated by the admissible control $\tilde{u}(\cdot) = u_k(\cdot)$. Observe that $u_k(\cdot)$ is required only on the final iteration of the iterative procedures. In some situations, however, it may be desirable to compute $u_k(\cdot)$ for each $k > 0$ in a recursive manner, using $u_{k-1}(\cdot)$ and those $u(\cdot, -z_i)$, $0 \leq i \leq k-1$ which occur on iteration $k-1$. If K is strictly convex, then $s(-z_k)$ will satisfy specified stopping criteria for k sufficiently large. Hence $\tilde{z}$ and $\tilde{u}(\cdot)$ can be taken to be $s(-z_k)$ and $u(\cdot, -z_k)$. This avoids the necessity for finding $u_k(\cdot)$, but examples show that an increased number of iterations are required.

There are several instances when it is possible to simplify the finding of an admissible $u_k(\cdot)$ which generates z_k. Consider IIP with Selection Rule D. Then z_k, $k > 0$, has a representation: $z_k = \sum_{i \in \hat{I}} \sigma_i s(-z_i)$

where $\hat{I}$ is a set of p + 1 or fewer integers from [0, k - 1], $\sigma_i \geq 0\ (i \in \hat{I})$ and $\sum_{i \in \hat{I}} \sigma_i = 1$. It follows that $u_k(\cdot)$ can be constructed from $u(t, -z_i)$, $i \in \hat{I}$. Since it may occur that $p + 1 << k - 1$ where k denotes the iteration on which the stopping criteria are satisfied, the use of Selection Rule D simplifies the construction of $\tilde{u}(\cdot) = u_k(\cdot)$. For example, $f(u, t) = B(t)u + b(t)$ and U convex imply $\tilde{u}(t) = u_k(t) = \sum_{i \in \hat{I}} \sigma_i u(t, -z_i)$. This same simplification is achieved in IIP with Selection Rule C if on some iteration prior to the final one, Phase C2 is entered. Furthermore, even with IIP (Selection Rules A or B) or BIP it may be possible to represent z_k as $\sum_{i \in \tilde{I}} \sigma_i s(-z_i)$ where $\tilde{I}$ is a proper subset of $\{0, 1, \ldots, k - 1\}$ (note: $\tilde{I}$ may contain more than p + 1 integers), $\sigma_i \geq 0\ (i \in \tilde{I})$, and $\sum_{i \in \tilde{I}} \sigma_i = 1$. This is the case if for some $j < k$ z_j has a zero coefficient in the convex combination expression for z_{j+1}.

Now consider Problem 2. Let $q(x)$, $x \in E^m$, be the quadratic form

$$q(x) = x \cdot Gx + g \cdot x\ , \tag{5.3.1}$$

where G and g are as in (5.1.6). It is easy to show that q(x) can be written as

$$q(x) = |\underline{G}x + \underline{g}|^2 + q_0, \tag{5.3.2}$$

where $\underline{G}$ is an $\ell \times m$ matrix, ℓ = rank G, $G = \underline{G}'\underline{G}$, $\underline{g} = \frac{1}{2}(\underline{G}\,\underline{G}')^{-1}\underline{G}g$ (that is, $g = 2\underline{G}'\underline{g}$), and $q_0 = -|\underline{g}|^2 = \min_{x \in E^m} q(x)$. Thus if K is defined by

$$K = \{z : z = \underline{G}x + \underline{g},\ x \in R(T)\}\ , \tag{5.3.3}$$

it follows that K is compact, convex, and

$$\min_{x \in R(T)} q(x) = \min_{z \in K} |z|^2 + q_0\ . \tag{5.3.4}$$

Consequently Problem 2 can be stated as a variant of BP: find an $x^* \in R(T)$ and an admissible control $u^*(\cdot)$ such that

$x^* \in \{x : \underline{G}x + \underline{g} = z^*,\ x \in R(T)\}$, $|z^*|^2 = \min_{z \in K} |z|^2$, and $x^* = x_{u^*}(T)$. Note that x^* is not necessarily unique. A contact function $s(y)$, $y \in E^n$, of K is

$$s(y) = \underline{G}s_R(T, \underline{G}'y) + \underline{g}. \tag{5.3.5}$$

This follows from

$$\max_{z \in K} y \cdot z = \max_{x \in R(T)} y \cdot (\underline{G}x + \underline{g}) = s_R(T, \underline{G}'y) \cdot \underline{G}'y + y \cdot g$$

$$= y \cdot (\underline{G}s_R(T, \underline{G}'y) + \underline{g}) . \tag{5.3.6}$$

Since for every $x \in R(T)$ there is a corresponding point $\underline{G}x + \underline{g} = z \in K$, BIP or IIP can be extended by simple substitution to Problem 2. For BIP Gilbert [Gl] has written out the complete details. The procedures yield a point $\tilde{x} \in R(T)$ such that $|\tilde{x} - x^*| < \epsilon$ (arbitrary $\epsilon > 0$), x^* an optimal state. An admissible control $\tilde{u}(\cdot)$ which generates $\tilde{x}$ is determined in a way similar to that described in Problem 1.

5.4 The Sets $K(\omega) = X(\omega) - Y(\omega)$

Before applying GIP to Problems 1 through 6, some results on the difference of two sets are stated. Let $\Omega = [0, \hat{\omega}]$, $\hat{\omega} > 0$, be a compact interval in E^1 and consider sets $X(\omega)$, $Y(\omega) \subset E^n$ which are compact, convex, and continuous on Ω. Define

$$K(\omega) = X(\omega) - Y(\omega)$$

$$= \{z : z = x - y,\ x \in X(\omega),\ y \in Y(\omega)\} . \tag{5.4.1}$$

It is easy to show that the sets $K(\omega)$ are compact, convex, and continuous on Ω. Furthermore, the support function $\eta(\omega, \cdot)$ of $K(\omega)$ can be written in terms of the support functions $\eta_X(\omega, \cdot)$ and $\eta_Y(\omega, \cdot)$ of $X(\omega)$ and $Y(\omega)$ as follows: (arbitrary $\bar{y} \in E^n$)

$$\eta(\omega, \bar{y}) = \max_{z \in K(\omega)} z \cdot \bar{y} = \max_{\substack{x \in X(\omega) \\ y \in Y(\omega)}} (x - y) \cdot \bar{y}$$

$$= \max_{x \in X(\omega)} x \cdot \bar{y} + \max_{y \in Y(\omega)} y \cdot (-\bar{y})$$

$$= \eta_X(\omega, \bar{y}) + \eta_Y(\omega, -\bar{y}) \ . \qquad (5.4.2)$$

From (5.4.2) it follows that if $s_X(\omega, \cdot)$ and $s_Y(\omega, \cdot)$ are contact functions of $X(\omega)$ and $Y(\omega)$, then a contact function of $K(\omega)$ is

$$s(\omega, \bar{y}) = s_X(\omega, \bar{y}) - s_Y(\omega, -\bar{y}), \ \bar{y} \in E^n. \qquad (5.4.3)$$

Clearly $0 \notin K(\omega)$ if and only if sets $X(\omega)$ and $Y(\omega)$ do not intersect.

5.5 Solution of Problems 1, 2, and 3 by GIP

Let $X(\omega)$, $\omega \in \Omega = [0, \hat{\omega}]$, be the fixed set $R(T)$ in each of Problems 1, 2, and 3. A reasonable choice for $\hat{\omega}$ is $\max_{x \in R(T)} \xi(x)$ where $\xi(x)$ equals $|x|$, $x \cdot Gx + g \cdot x$, $h^0(x)$ in Problems 1, 2, 3 respectively.

Define $Y(\omega)$, $\omega \in \Omega$, as follows:

$$Y(\omega) = \{x \in E^m : |x| \le \omega\} \text{ for Problem 1,} \qquad (5.5.1)$$

$$Y(\omega) = \{x \in E^m : x \cdot Gx + g \cdot x \le \omega\} \text{ for Problem 2,} \qquad (5.5.2)$$

$$Y(\omega) = \{x \in E^m : h^0(x) \le \omega\} \text{ for Problem 3.} \qquad (5.5.3)$$

Since $|x|$, $x \cdot Gx + g \cdot x$, and $h^0(x)$ are convex functions on E^m it follows that these sets $Y(\omega)$ are convex and continuous on Ω. Clearly $Y(\omega)$ in (5.5.1) is compact, and if it is assumed that G is non-singular then $Y(\omega)$ in (5.5.2) is compact. The compactness of $\{x : h^0(x) \le \text{constant}\}$ was assumed in (5.1.7). (Methods exist for treating convex functions $h^0(\cdot)$ which do not satisfy this assumption.) Thus if $K(\omega) = X(\omega) - Y(\omega)$, $K(\omega)$ is compact, convex, and continuous on Ω for Problems 1, 2, 3. Consequently each of these problems can be stated as a variant of GP: find $\omega^* \in \Omega$, a point $x^* \in R(T)$, and an admissible control $u^*(\cdot)$ such that

$0 \notin K(\omega)$, $0 \leq \omega < \omega^*$, $0 \in K(\omega^*)$, $x^* \in Y(\omega^*)$, and $x^* = x_{u^*}(T)$. The quantity ω^* is the minimum cost in each problem.

To apply GIP it is necessary that a contact function $s(\omega, y)$ of $K(\omega)$ be available. Since $X(\omega) = R(T)$, $s_X(\omega, y) = s_R(T, y)$ and thus (5.4.3) yields $s(\omega, y)$ if $s_Y(\omega, y)$ can be determined. In Problem 1 $s_Y(\omega, y) = \omega |y|^{-1} y$ for $|y| > 0$; $s_Y(\omega, y) =$ arbitrary $x \in Y(\omega)$ for $|y| = 0$. Finding $s_Y(\omega, y)$ in Problems 2 and 3 is more difficult. Consider Problem 2. Since the gradient of $x \cdot Gx + g \cdot x$ is $2Gx + g$, $s_Y(\omega, y)$ must satisfy (for $|y| > 0$):

$$2Gs_Y(\omega, y) + g = \mu y, \quad \mu > 0 \tag{5.5.4}$$

and

$$s_Y(\omega, y) \cdot Gs_Y(\omega, y) + g \cdot s_Y(\omega, y) = \omega. \tag{5.5.5}$$

These are a set of $m + 1$ equations in $m + 1$ unknowns: $s_Y(\omega, y)$ and μ. Any solution for $s_Y(\omega, y)$ will suffice but it may not be easy to determine. Similarly in Problem 3 if the gradiant of $h^0(x)$ exists, it follows that $s_Y(\omega, y)$ must satisfy (for $|y| > 0$):

$$\text{grad } h^0(s_Y(\omega, y)) = \mu y, \quad \mu > 0 \tag{5.5.6}$$

and

$$h^0(s_Y(\omega, y)) = \omega. \tag{5.5.7}$$

For convergence of GIP it is required that the support function $\eta(\omega, y)$ of $K(\omega)$ satisfy: $\dfrac{\eta(\omega_a, y) - \eta(\omega_b, y)}{\omega_a - \omega_b}$ bounded from above for all ω_a, $\omega_b \in \Omega$, $\omega_a \neq \omega_b$, bounded by $y \in E^m$. Since $X(\omega)$ does not vary with ω in Problems 1, 2, 3, (5.4.2) implies that this requirement is satisfied if $\dfrac{\eta_Y(\omega_a, y) - \eta_Y(\omega_b, y)}{\omega_a - \omega_b}$ is bounded from above (ω_a, ω_b, y as before). For Problem 1 $\eta_Y(\omega, y) = y \cdot s_Y(\omega, y) = \omega |y|$, $|y| > 0$, so the desired condition is true.

Assuming that a contact function $s(\omega, y)$ is available and that $\eta(\omega, y)$ satisifes the requirement of the preceding paragraph, GIP may be used to generate $\tilde{\omega} \in [0, \omega^*]$ and a point $\tilde{z} \in K(\tilde{\omega})$ such that $\tilde{z} = \tilde{x} - \tilde{y}$, $\tilde{x} \in X(\tilde{\omega})$, $\tilde{y} \in Y(\tilde{\omega})$, where $|\tilde{z}| < \epsilon$ and $\min_{y \in Y(\omega^*)} |\tilde{x} - y| < \epsilon$ (arbitrary $\epsilon > 0$). For Problems 1 and 2 $|\tilde{x} - x^*| < \epsilon$, where x^* is an optimal state (x^* unique for Problem 1). An admissible control $\tilde{u}(\cdot)$ which will generate $\tilde{x}$ can be found by a method similar to that described in Section 5.3 since $\tilde{x}$ can be written as a convex combination of an initial point $x_0 \in R(T)$ and a certain number of contact points of $R(T)$.

5.6 Solution of Problems 4, 5, and 6 by GIP

Consider Problem 4 and choose $\hat{\omega} = \underset{(x^0, x) \in \underline{R}(t)}{\text{maximum}} x^0$, where $\underline{R}(t)$ is defined by (5.2.18). The function $f^0(\cdot, \cdot)$ in (5.1.8) may be chosen so that the lower boundary of the closed interval Ω is 0. For $\omega \in \Omega = [0, \hat{\omega}]$, let $X(\omega) = \underline{R}(T)$ and $Y(\omega) = \{(\omega, w) : w \in W(T)\}$. Since $W(T)$ is compact and convex, $Y(\omega)$ has these properties for each $\omega \in \Omega$. The fact that $Y(\omega)$ is continuous on Ω is obvious. Hence $K(\omega) = X(\omega) - Y(\omega)$ is compact, convex, and continuous on Ω and Problem 4 is a variant of GP: find $\omega^* \in \Omega$, a point $\underline{x}^* \in \underline{R}(T)$, and admissible control $u^*(\cdot)$ such that $0 \notin K(\omega)$, $0 \leq \omega < \omega^*$, $0 \in K(\omega^*)$, $\underline{x}^* \in Y(\omega)$, and $\underline{x}^* = \underline{x}_{u^*}(T)$. The minimum cost (fuel) is ω^*.

A contact function $s_X(\omega, \underline{y})$, $y \in E^{m+1}$, is $s_{\underline{R}}(T, \underline{y})$ given by (5.2.20). Suppose a contact function $s_W(y)$, $y \in E^m$, of $W(T)$ is known. Write $\underline{y} = (y^0, y)$ where $y^0 \in E^1$, $y \in E^m$. Then a contact function $s_Y(\omega, \underline{y})$ of $Y(\omega)$ is $(\omega, s_W(y))$ for $|y| > 0$; $s_Y(\omega, \underline{y})$ = arbitrary $\underline{x} \in Y(\omega)$ for $|y| = 0$. If $W(T)$ is a single point $w(T)$, then $s_Y(\omega, \underline{y}) = (\omega, w(T))$ for all $\underline{y}$. A contact function of $K(\omega)$ is $s(\omega, \underline{y}) = s_X(\omega, \underline{y}) - s_Y(\omega, -\underline{y})$. Since $\eta_X(\omega, \underline{y})$ does not vary with ω and $\eta_Y(\omega, \underline{y}) = \underline{y} \cdot s_Y(\omega, \underline{y})$, it is clear that the condition on the difference quotient of $\eta(\omega, \underline{y})$ in Theorem 4.3.1 is satisfied.

For Problem 5 define sets

$$U_\lambda = \{u : |u^i| \le \lambda, \quad i = 1, 2, \ldots, r\}, \tag{5.6.1}$$

and

$$C_\lambda(T) = \{ \int_0^T \Phi^{-1}(\sigma) B(\sigma) u(\sigma) d\sigma : u(\cdot) \text{ measurable with range in } U_\lambda \} \tag{5.6.2}$$

A control $u(\cdot)$ on $[0, T]$ is said to have effort λ if $\sup_{\substack{0 \le t \le T \\ 1 \le i \le r}} |u^i(t)| = \lambda$. Suppose $\Phi^{-1}(T) w(T) - x(0) \in C_\lambda(T)$. Then there exists a measurable $u(\cdot)$ with range in U_λ such that $w(T) = \Phi(T) x(0) + \int_0^T \Phi(T)\Phi^{-1}(\sigma) B(\sigma) u(\sigma) d\sigma$, which implies that $u(\cdot)$ transfers x from $x(0)$ to $w(T)$ in time T. Observe $C_\lambda(T) = \lambda C_1(T) = \{\bar{c} : \bar{c} = \lambda c, \ c \in C_1(T)\}$. It follows that if λ^* is the smallest λ such that $\Phi^{-1}(T)w(T) - x(0) \in C_\lambda(T)$ and ω^* is the largest ω such that $\omega[\Phi^{-1}(T) w(T) - x(0)] \in C_1(T)$, then $\lambda^* = (\omega^*)^{-1}$.

One choice for $\hat{\omega}$ is $|\Phi^{-1}(T) w(T) - x(0)|^{-1} \max_{c \in C_1(T)} |c|$. For $\omega \in \Omega = [0, \hat{\omega}]$ let $X(\omega) = C_1(T)$ and let $Y(\omega)$ be the point $\omega[\Phi^{-1}(T)w(T) - x(0)]$. Since $C_1(T)$ is compact and convex (Section 5.2), $K(\omega) = X(\omega) - Y(\omega)$ has these properties. Clearly $K(\omega)$ is continuous on Ω and has a contact function $s(\omega, y) = s_{C_1}(T, y) - \omega[\Phi^{-1}(T) w(T) - x(0)]$, where $s_{C_1}(T, y)$ is given by (5.2.15) and (5.2.13) with $U = U_1$ and $f(u, t) = B(t)u$. Thus Problem 5 is a variant of GP: find $\omega^* \in \Omega$, $\omega^* > 0$, and a measurable control $u^*(\cdot)$ with range in U_1 such that $0 \notin K(\omega)$, $0 \le \omega < \omega^*$, $0 \in K(\omega^*)$, and $x_{(\omega^*)^{-1} u^*}(T) = w(T)$. The function $(\omega^*)^{-1} u^*(\cdot)$ is a minimum effort control and $(\omega^*)^{-1}$ is the minimum effort. Note that the support function $\eta(\omega, y) = y \cdot s_{C_1}(T, y) - \omega y \cdot [\Phi^{-1}(T) w(T) - x(0)]$ has a difference quotient which satisfies the requirement in Theorem 4.3.1.

Now consider Problem 6. Let $\omega = t$, $\Omega = \Theta$, $X(\omega) = R(t)$, and $Y(\omega) = W(t)$. Then $K(\omega) = X(\omega) - Y(\omega)$ is compact, convex, and continuous on Ω. A contact function of $K(\omega)$ is $s(\omega, y) = s_R(\omega, y) - s_W(\omega, y)$, $y \in E^m$, where $s_R(\cdot, \cdot)$ is given by (5.2.8) and $s_W(\omega, \cdot)$ is a contact function of $W(\omega)$.

The optimization objective in Problem 6 is: find $\omega^* \in \Omega$, a point $x^* \in R(\omega^*)$, and an admissible control $u^*(\cdot)$ such that $0 \notin K(\omega)$, $0 \le \omega < \omega^*$, $0 \in K(\omega^*)$, $x^* \in W(\omega^*)$, and $x^* = x_{u^*}(\omega^*)$. The minimum time is ω^*.

The support function of $K(\omega)$ is $\eta(\omega, y) = \eta_R(\omega, y) + \eta_W(\omega, -y)$. Assume that $\dfrac{\eta_W(\omega_a, y) - \eta_W(\omega_b, y)}{\omega_a - \omega_b}$ is bounded from above for all ω_a, $\omega_b \in \Omega$, $\omega_a \neq \omega_b$, bounded $y \in E^m$. This is certainly true if the target set $W(\omega)$ is a single point $w(\omega)$ which satisfies a Lipschitz condition on Ω. From (5.2.6)

$$\eta_R(\omega, y) = y \cdot \Phi(\omega)\, x(0) + \int_0^{\omega} \max_{v \in U} [\psi(\sigma, y) \cdot f(v, \sigma)]\, d\sigma \,. \qquad (5.6.3)$$

Thus

$$\frac{\partial \eta_R(\omega, y)}{\partial \omega} = y \cdot A(\omega)\, \Phi(\omega)\, x(0) + \max_{v \in U} [\psi(\omega, y) \cdot f(v, \omega)]. \qquad (5.6.4)$$

Since $A(\cdot)$, $\Phi(\cdot)$, $\psi(\cdot, y)$, and $f(v, \cdot)$ are continuous on Ω, $\dfrac{\partial \eta_R(\omega, y)}{\partial \omega}$ exists and is bounded for all bounded y. It follows that $\dfrac{\eta_R(\omega_a, y) - \eta_R(\omega_b, y)}{\omega_a - \omega_b}$ and $\dfrac{\eta(\omega_a, y) - \eta(\omega_b, y)}{\omega_a - \omega_b}$ are bounded from above for all ω_a, $\omega_b \in \Omega$, $\omega_a \neq \omega_b$, bounded $y \in E^m$.

Consequently for Problems 4, 5, and 6 GIP can be applied and convergence is guaranteed. An arbitrary admissible control is chosen to initiate the procedure. Then GIP is continued until a point $\tilde{z} \in K(\tilde{\omega})$, $\tilde{\omega} \in \Omega$, satisfying specified terminal criteria is found. Finally an admissible control $\tilde{u}(\cdot)$ which generates an approximate optimal state is determined in a manner similar to that described in Section 5.3.

5.7 Numerical Results for GIP Applied to a Minimum-Fuel Example

Barr and Gilbert [B3] presented an iterative technique for solving minimum-fuel terminal control problems. The technique is a special case of GIP and uses BIP for the minimization problems which occur in Step 1 of each iteration. Numerical results for a minimum-fuel example using this technique were obtained by Hutcheson [H5]. Computations have been carried out for this same example using IIP (with Selection Rule A and p = n) in Step 1 of each iteration. Results from these computations are presented in this section and are compared with Hutcheson's data.

The example treated is a particular case of Problem 4: $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $f = \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$, $U = \{u : |u| \leq 1\}$, $a = 0$, $f^0 = |u|$, $T = 4$, $W(T) = 0$. Because of the simplicity of this example, analytical results can be obtained [e.g., F5]. In particular, the set of all initial states x(0) which can be transferred to the origin in time T with admissible controls can be drawn in the 2-dimensional state space. This set is shown in Figure 5.7.1, which also indicates some isofuel contours and the initial states used for the numerical computations. Clearly the minimum fuel $\omega^* \in [0, 4]$ and if $x(0) \neq 0$, $\omega^* > 0$. For an initial state x(0) lying in the shaded region there is no unique optimal control. Since the isofuel contours contain straight-line segments, it follows that the reachable set $\underline{R}(T) \subset E^3$ is not strictly convex. Thus the gradient-type convexity methods are not applicable to the example.

A detailed description of the iterative procedure for the case when BIP is used in Step 1 of GIP is presented in [B3]. The procedure is much the same when IIP is used. For this simple example it is not necessary to actually solve (5.2.17), (5.2.19), and (5.2.21) on the computer. As described by Hutcheson, a contact function $s_{\underline{R}}(T, \underline{y})$ can be evaluated in terms of certain switching times which are related algebraically to the vector $\underline{y}$.

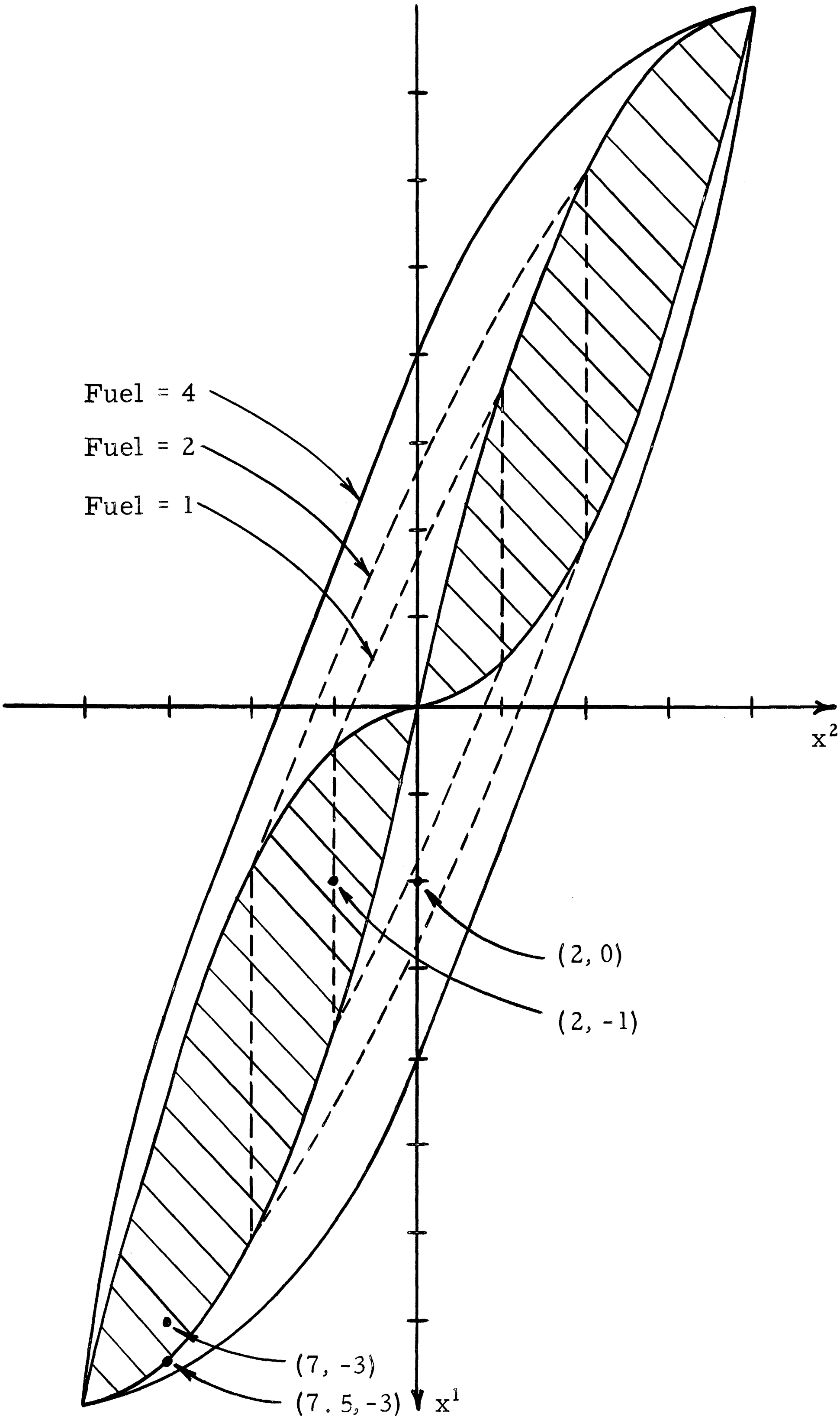


Figure 5.7.1 Initial conditions and isofuel contours for the minimum-fuel example.

Let e be the unit vector (1, 0, 0). Then according to Section 5.6 $K(\omega)$, $\omega \in \Omega = [0, 4]$, is equal to $\underline{R}(T) - \omega e$. For $\underline{x} \in \underline{R}(T)$ define the scalar functions

$$\overline{\rho}(\underline{x}) = |\underline{x} - (\underline{x} \cdot e)e|, \tag{5.7.1}$$

$$\overline{\phi}(\underline{x}) = \underline{x} \cdot e . \tag{5.7.2}$$

If $\underline{x} = (x^0, x)$, $\overline{\rho}(\underline{x})$ is the Euclidean distance of the state x from the origin and $\overline{\phi}(\underline{x}) = x^0$ is the fuel associated with the state x. Suppose an $\omega \in [0, \omega^*]$ and a $z \in K(\omega)$ are generated by GIP. Since $z + \omega e \in \underline{R}(T)$ and $\overline{\phi}(z + \omega e) - \omega^* < \overline{\phi}(z + \omega e) - \omega$, it follows that $\overline{\phi}(z + \omega e) - \omega$ is an upper bound on the difference between the fuel corresponding to $z = (x^0, x) - \omega e$ and the minimum fuel ω^*; furthermore, $\overline{\rho}(z + \omega e)$ measures the error in satisfying the terminal constraint $x = 0$. Hence a reasonable computational objective is to find $z \in K(\omega)$ such that $\overline{\phi}(z + \omega e) - \omega \leq \epsilon_1$ and $\overline{\rho}(z + \omega e) \leq \epsilon_2$ (ϵ_1, ϵ_2 specified arbitrary positive constants).

Consider GIP. Let $\ell(i)$ denote the number of points generated by BIP or IIP (including the initial point) which occur on iteration i of GIP. The kth point generated by BIP or IIP on iteration i of GIP will be denoted z(j) where

$$j = k + \sum_{q=0}^{i=1} \ell(q), \quad i \neq 0, \quad k = 0, 1, \ldots, \ell(i) - 1 \tag{5.7.3}$$

$$= k, \qquad i = 0, \quad k = 0, 1, \ldots, \ell(0) - 1.$$

This notation avoids the use of a double index. In the notation of Chapter 4 then, $z_i = z(\sum_{q=0}^{i} \ell(q) - 1)$. Note that j + 1 represents the number of points generated up to and including z(j). Also j is the number of times the contact function has been evaluated. Define

$$\overline{\rho}_j = \overline{\rho}(z(j) + \omega_i e), \tag{5.7.4}$$

$$\overline{\phi}_j = \overline{\phi}(z(j) + \omega_i e), \tag{5.7.5}$$

and

$$\bar{\delta}_j = \bar{\phi}(z(j) + \omega_i e) - \omega_i. \tag{5.7.6}$$

The sequences $\{\bar{\rho}_j\}$, $\{\bar{\phi}_j\}$, and $\{\bar{\delta}_j\}$ are used to display the numerical results.

A fixed $\theta_i = \theta$ was used for each computation and various initial states $x(0)$ (shown in Figure 5.7.1) were investigated. For iteration 0 of GIP a point $z(0)$ in $K(0) = \underline{R}(T)$ with which to initiate BIP or IIP was found by solving (5.2.17) analytically using the admissible control $u(t) = \frac{1}{2}$, $t \in [0, 4]$. For iteration $i > 0$ of GIP the point $z_{i-1} - (\omega_i - \omega_{i-1})e$ was used to initiate BIP or IIP (see Figure 5.7.2).

In Table 5.7.1 the first j for which $\bar{\delta}_j \leq \epsilon$ and $\bar{\rho}_j \leq \epsilon$ is shown for the case $x(0) = (2, -1)$ and a variety of θ values. The same quantities are shown in Tables 5.7.2, 5.7.3, and 5.7.4 for the cases $x(0) = (2, 0)$, $x(0) = (7, -3)$, and $x(0) = (7.5, -3)$. A comparison of the data in A and B of Tables 5.7.1 and 5.7.2 makes clear the significant improvement achieved when IIP is used rather than BIP. Observe that there is little dependence on θ; however, an intermediate value, say $\theta = .4$, seems to be most desirable.

Figures 5.7.3 through 5.7.6 show the details of the sequences $\{\bar{\delta}_j\}$ and $\{\bar{\rho}_j\}$ for $\theta = .4$, BIP and IIP, and initial states (2, 0) and (2, -1). Similar behavior was observed for other θ values.

Note the especially significant improvement which was achieved for the initial state (2, -1), which lies in the shaded region of Figure 5.7.1. That convergence is indeed rapid for initial states in this region is verified by the data for $x(0) = (7, -3)$ and $x(0) = (7.5, -3)$ in Tables 5.7.3 and 5.7.4.

Tables 5.7.5 through 5.7.10 show the functions ω_i and $\ell(i)$ for the various θ, initial states, and BIP and IIP. Furthermore, the sequence $\{\bar{\phi}_j\}$ is illustrated in Figures 5.7.7 and 5.7.8. This again shows the

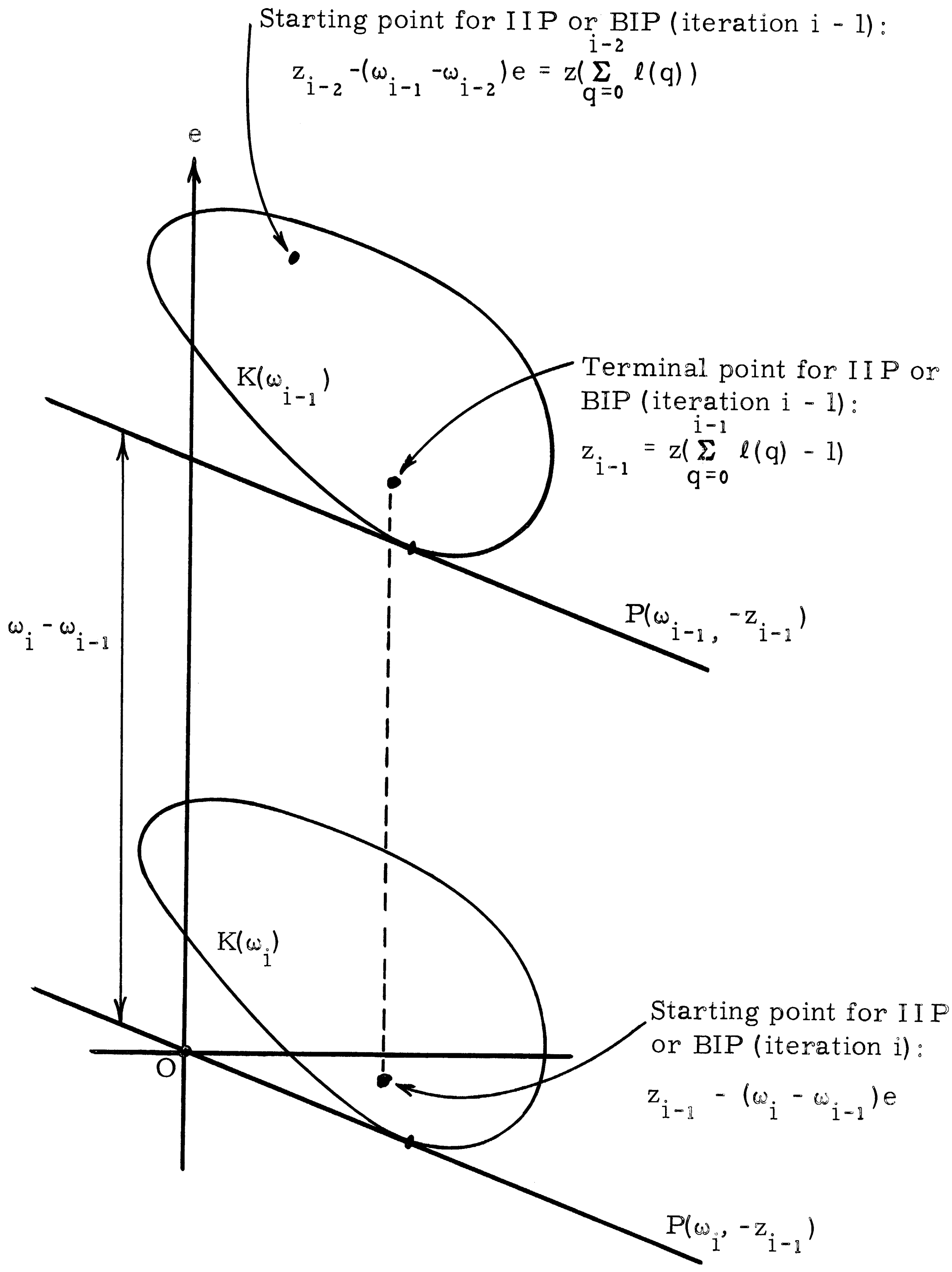


Figure 5.7.2 Geometry for minimum-fuel example.

improvement possible when IIP is used rather than BIP on each iteration of GIP.

It can be seen that a sizeable decrease in $\bar{\delta}_j$ occurs on transitions where i is increased by one (i. e., transitions from $k \neq 0$ to $k = 0$). On these transitions both $\bar{\rho}_j$ and $\bar{\phi}_j$ are constant.

Table 5.7.1 The first j for which $\bar{\delta}_j \leq \epsilon$ and $\bar{\rho}_j \leq \epsilon$; x(0) = (2, -1).

	A						B					
θ \ ε	1	.5	.1	.05	.01	10^{-3}	1	.5	.1	.05	.01	10^{-3}
.1	3	4	11	11	11	11	4	8	199	>499		
.3	3	4	11	11	11	11	4	8	153	>499		
.4	3	4	11	11	11	11	4	8	85	>1499		
.5	3	4	11	11	11	11	4	8	81	>499		
.7	3	5	9	9	9	9	4	8	>199			
.9	3	6	7	11	11	11	4	28	>199			

A: IIP Selection Rule A with p = 3 used on each iteration of GIP.
B: BIP used on each iteration of GIP.

Table 5.7.2 The first j for which $\bar{\delta}_j \leq \epsilon$ and $\bar{\rho}_j \leq \epsilon$; x(0) = (2, 0).

	A						B					
θ \ ε	1	.5	.1	.05	.01	10^{-3}	1	.5	.1	.05	.01	10^{-3}
.1	4	5	16	19	29	39	4	7	94	96	379	>499
.3	2	5	12	15	21	27	4	9	76	91	244	>499
.4	2	5	14	14	14	24	3	13	25	40	129	436
.5	2	5	14	14	14	24	3	13	25	40	349	>499
.7	2	5	13	21	31	45	3	12	70	139	295	>499
.9	2	7	18	24	34	41	3	17	85	99	>99	

A: IIP Selection Rule A with p = 3 used on each iteration of GIP.
B: BIP used on each iteration of GIP.

Table 5.7.3 The first j for which $\bar{\delta}_j \le \epsilon$ and $\bar{\rho}_j \le \epsilon$; $x(0) = (7, -3)$; IIP Selection Rule A with p = 3 used on each iteration of GIP.

θ \ ε	1	.5	.1	.05	.01	10^{-3}
.1	2	4	11	11	11	11
.3	2	4	11	11	11	11
.4	2	4	11	11	11	11
.5	2	4	11	11	11	11
.7	2	5	7	7	27	27
.9	4	6	14	15	15	15

Table 5.7.4 The first j for which $\bar{\delta}_j \le \epsilon$ and $\bar{\rho}_j \le \epsilon$; $x(0) = (7.5, -3)$; IIP Selection Rule A with p = 3 used on each iteration of GIP.

θ \ ε	1	.5	.1	.05	.01	10^{-3}
.1	2	5	9	9	11	16
.3	2	3	8	8	8	8
.4	2	3	8	8	8	8
.5	2	3	8	8	8	8
.7	2	3	8	8	8	8
.9	4	6	13	16	21	26

Table 5.7.5 The functions ω_i and $\ell(i)$ for $x(0) = (2, -1)$; IIP Selection Rule A with $p = 3$ used on each iteration of GIP.

	θ \ i	0	1	2
ω_i	.1	0	.527	1.000
	.3	0	.527	1.000
	.4	0	.527	1.000
	.5	0	.527	1.000
	.7	0	.659	1.000
	.9	0	.852	1.000
$\ell(i)$	.1	4	2	5
	.3	4	2	5
	.4	4	2	5
	.5	4	2	5
	.7	5	2	2
	.9	6	3	2

Table 5.7.6 The functions ω_i and $\ell(i)$ for $x(0) = (2, -1)$; BIP used on each iteration of GIP.

	θ \ i	0	1	2	3	4	5	6	7	8
ω_i	.1	0	.599	.725	.793	.857	.908	.935	.954	.968
	.3	0	.599	.806	.928					
	.4	0	.599	.823	.945					
	.5	0	.776	.946						
	.7	0	.776							
	.9	0	.845							
$\ell(i)$	.1	3	9	4	4	2	48	129	260	>40
	.3	3	14	2	>480					
	.4	4	16	62	>1417					
	.5	7	68	>424						
	.7	7	>193							
	.9	27	>172							

Table 5.7.7 The functions ω_i and $\ell(i)$ for $x(0) = (2, 0)$; IIP Selection Rule A with $p = 3$ used on each iteration of GIP.

	θ \ i	0	1	2	3	4	5
ω_i	.1	0	.545	1.032	1.137	1.161	1.171
	.3	0	1.000	1.101	1.168	1.171	
	.4	0	1.000	1.169	1.171		
	.5	0	1.000	1.169	1.171		
	.7	0	1.000	1.152	1.171		
	.9	0	1.071	1.169	1.171		
$\ell(i)$	.1	2	5	5	10	4	13
	.3	4	7	2	13	1	
	.4	4	9	10	1		
	.5	4	9	10	1		
	.7	4	10	18	13		
	.9	5	21	14	1		

Table 5.7.8 The functions ω_i and $\ell(i)$ for $x(0) = (2, 0)$; BIP used on each iteration of GIP.

	θ \ i	0	1	2	3	4	5
ω_i	.1	0	.545	1.113	1.150	1.166	1.171
	.3	0	.560	1.138	1.171		
	.4	0	.897	1.126	1.167	1.171	
	.5	0	.897	1.126	1.171		
	.7	0	.907	1.117	1.167		
	.9	0	.961	1.166			
$\ell(i)$	.1	1	4	74	26	45	>349
	.3	3	2	96	>398		
	.4	7	15	22	127	265	
	.5	7	15	80	>397		
	.7	9	57	99	>334		
	.9	15	67	>17			

Table 5.7.9 The functions ω_i and $\ell(i)$ for $x(0) = (7, -3)$; IIP Selection Rule A with $p = 3$ used on each iteration of GIP.

	θ \ i	0	1	2	3
ω_i	.1	0	1.750	2.647	3.000
	.3	0	1.750	2.647	3.000
	.4	0	1.750	2.647	3.000
	.5	0	1.750	2.647	3.000
	.7	0	2.002	3.000	
	.9	0	2.543	3.000	
$\ell(i)$	.1	1	2	7	1
	.3	1	2	7	1
	.4	1	2	7	1
	.5	1	2	7	1
	.7	2	4	21	
	.9	3	7	5	

Table 5.7.10 The functions ω_i and $\ell(i)$ for $x(0) = (7.5, -3)$; IIP Selection Rule A with $p = 3$ used on each iteration of GIP.

	θ \ i	0	1	2	3
ω_i	.1	0	2.125	2.545	3.000
	.3	0	2.125	3.000	
	.4	0	2.125	3.000	
	.5	0	2.125	3.000	
	.7	0	2.125	3.000	
	.9	0	3.000		
$\ell(i)$	.1	1	2	3	10
	.3	1	3	4	
	.4	1	3	4	
	.5	1	3	4	
	.7	1	3	4	
	.9	3	23		

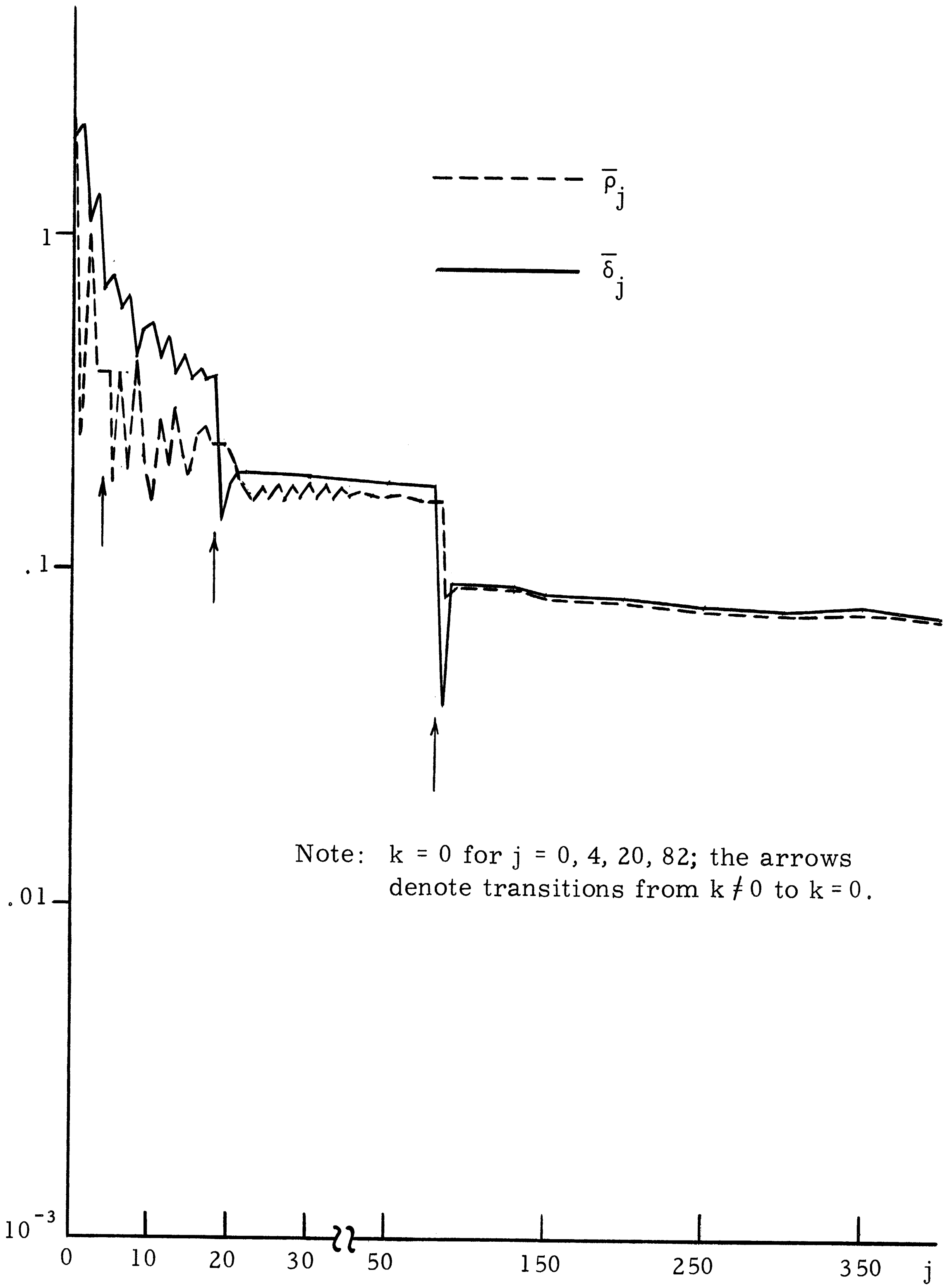


Figure 5.7.3 $\overline{\delta}_j$ and $\overline{\rho}_j$ for $x(0) = (2, -1)$, $\theta = .4$; BIP used on each iteration of GIP.

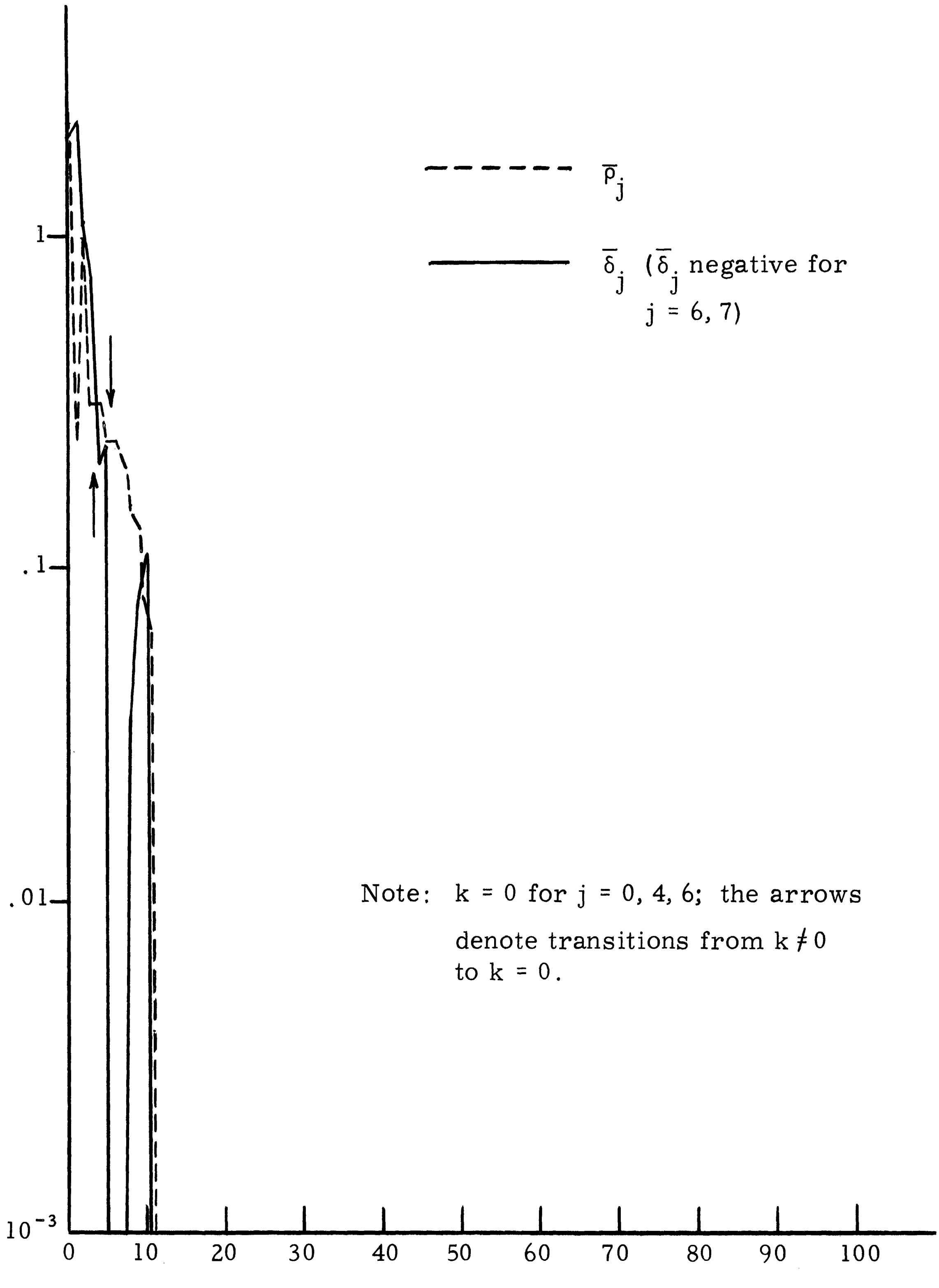


Figure 5.7.4 $\bar{\delta}_j$ and $\bar{\rho}_j$ for $x(0) = (2, -1)$, $\theta = .4$; IIP Selection Rule A with $p = 3$ used on each iteration of GIP.

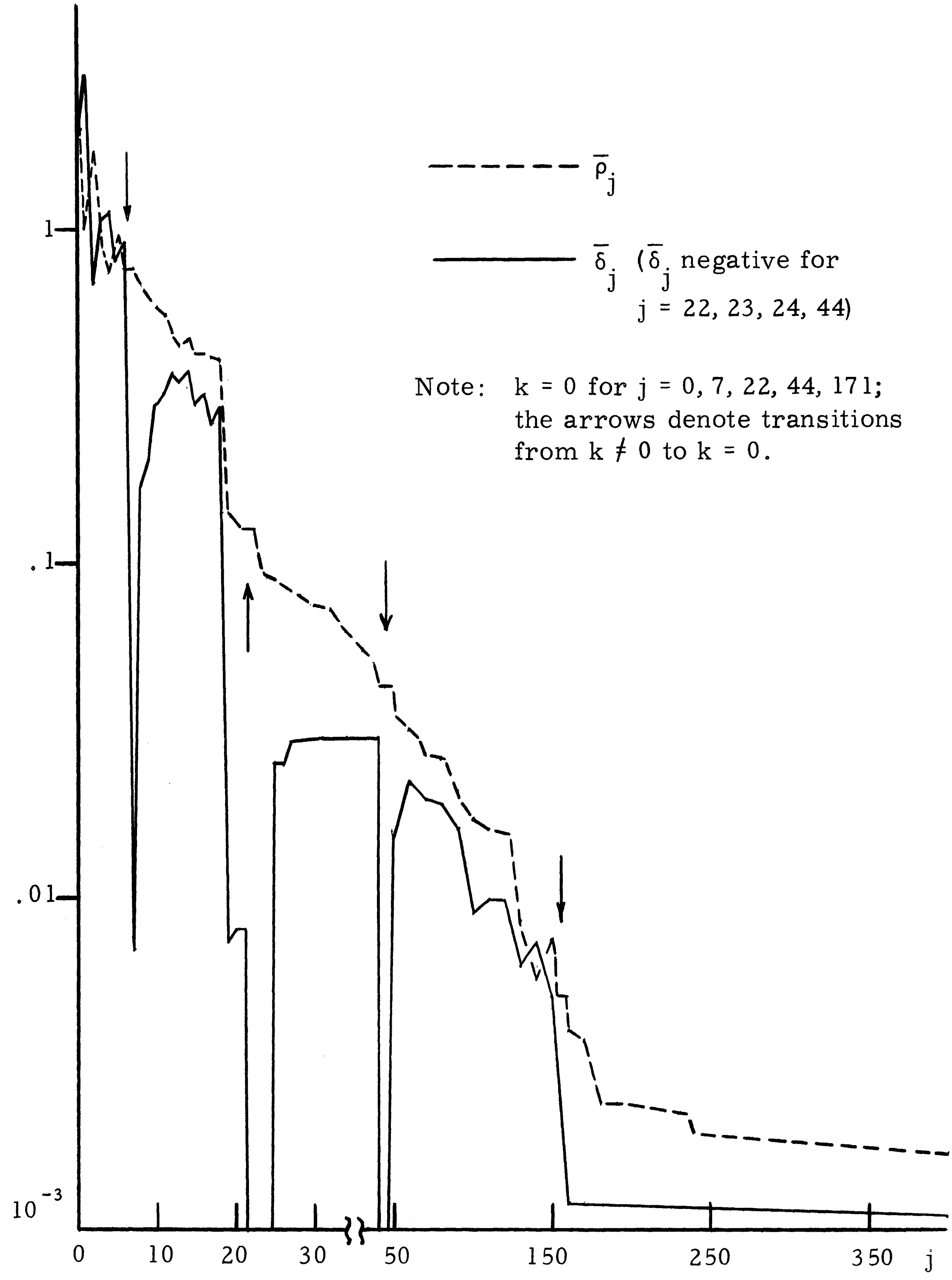


Figure 5.7.5 $\bar{\delta}_j$ and $\bar{\rho}_j$ for x(0) = (2, 0), θ = .4; BIP used on each iteration of GIP.

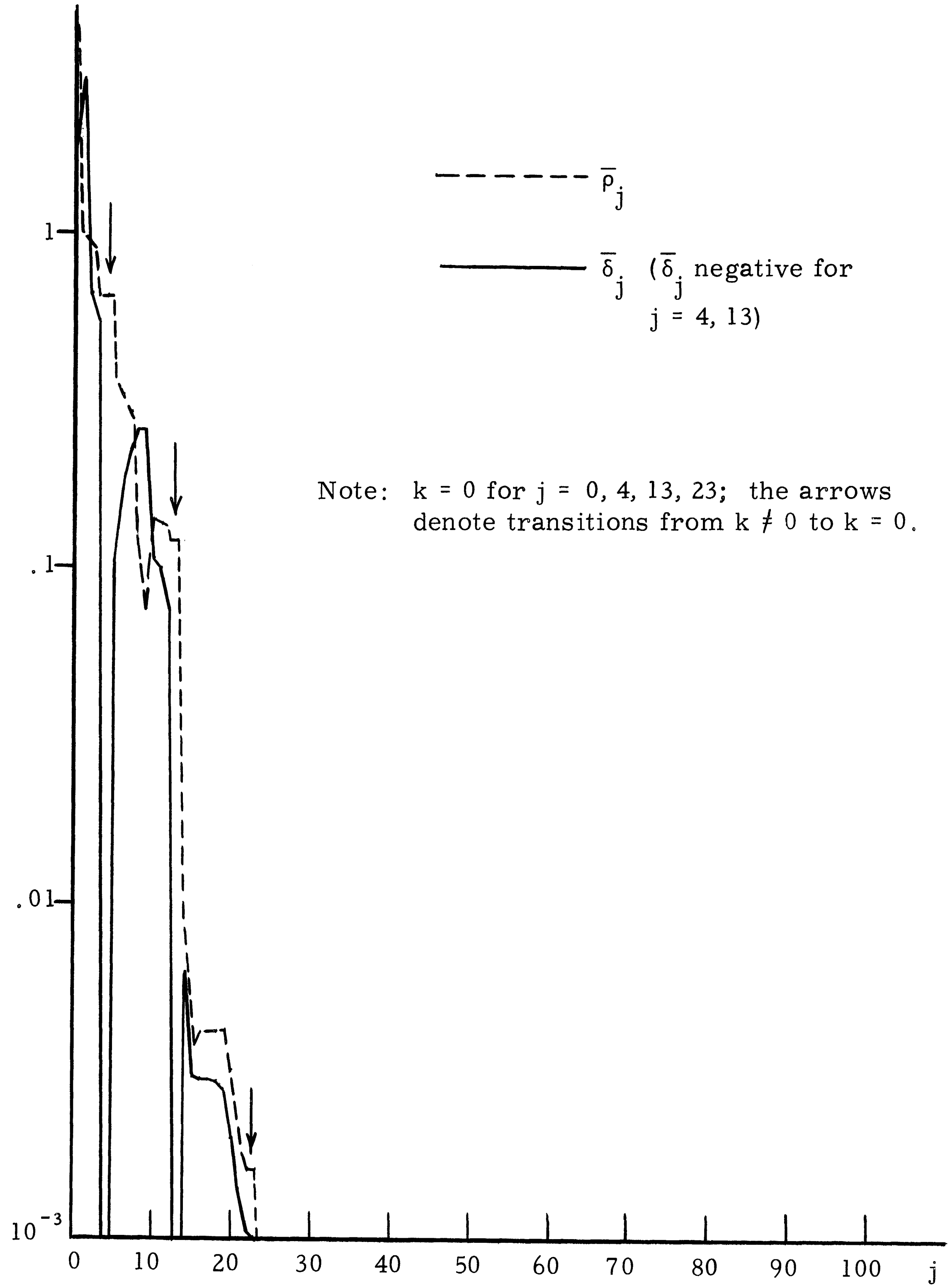


Figure 5.7.6 $\bar{\delta}_j$ and $\bar{\rho}_j$ for $x(0) = (2, 0)$, $\theta = .4$; IIP Selection Rule A with $p = 3$ used on each iteration of GIP.

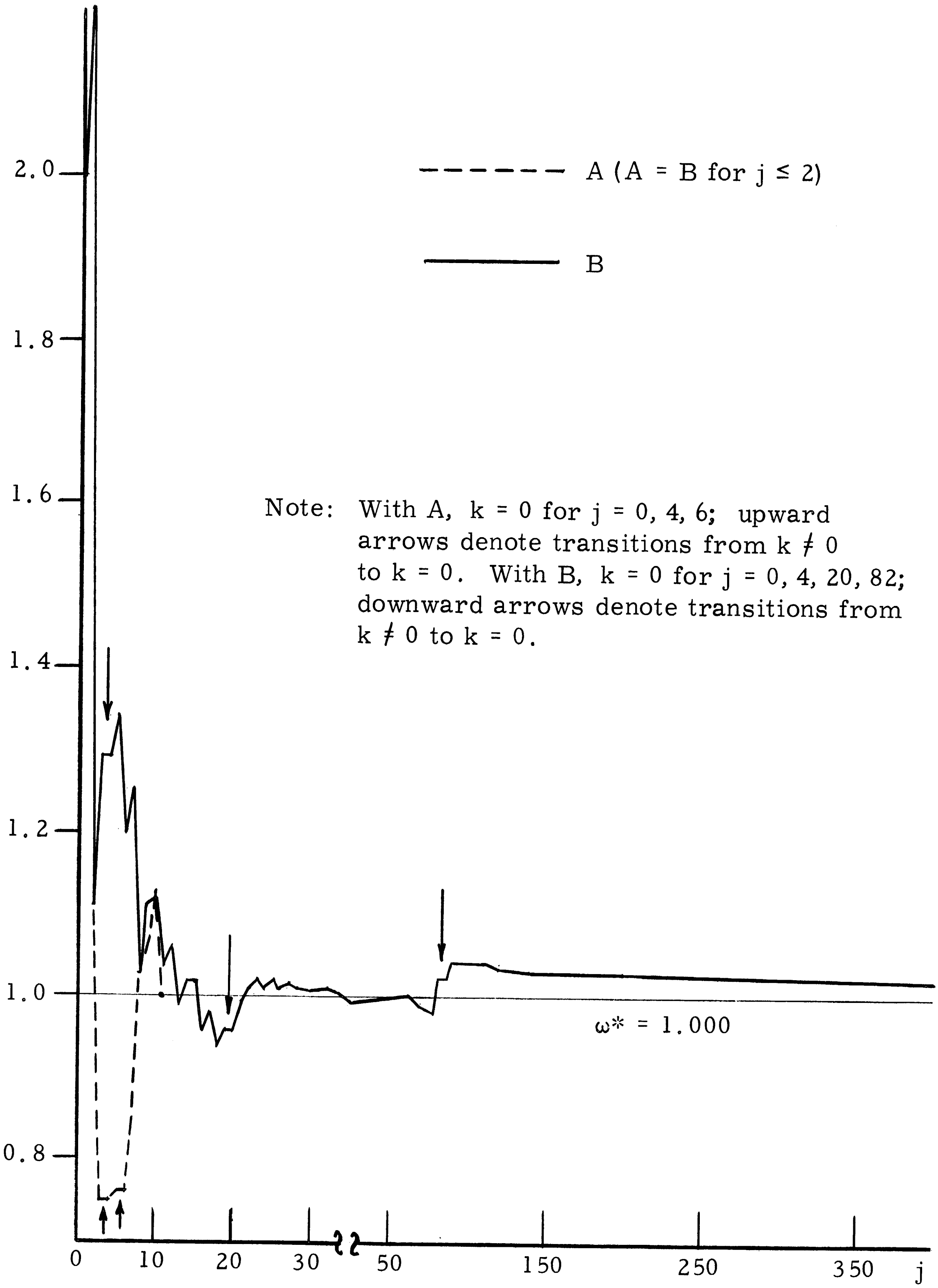


Figure 5.7.7 $\bar{\phi}_j$ for x(0) = (2, -1), θ = .4: A) IIP Selection Rule A with p = 3 used on each iteration of GIP; B) BIP used on each iteration of GIP.

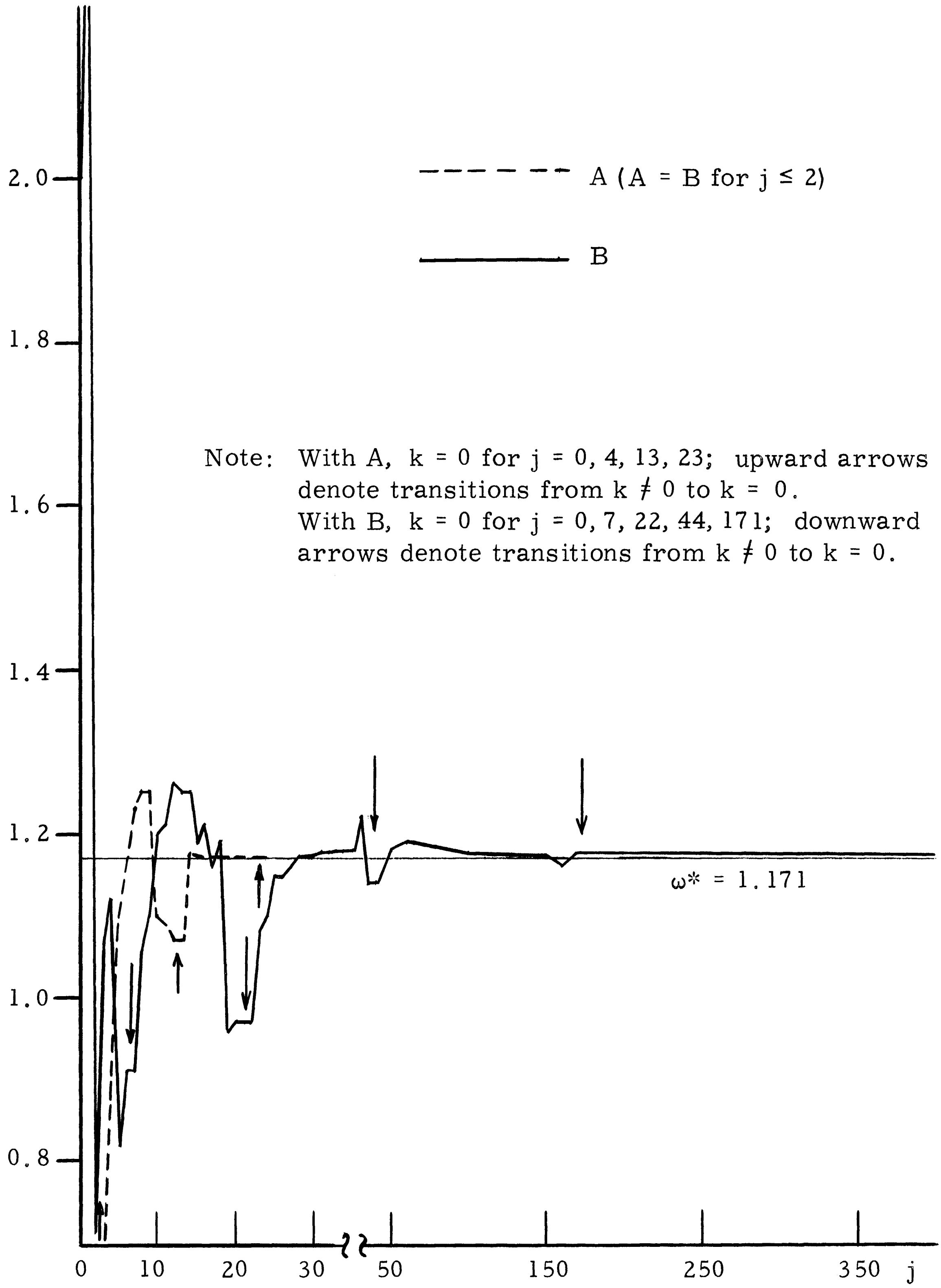


Figure 5.7.8 $\bar{\phi}_j$ for x(0) = (2, 0), θ = .4: A) IIP Selection Rule A with p = 3 used on each iteration of GIP; B) BIP used on each iteration of GIP.

List of Symbols

$A(t)$	continuous $m \times m$ matrix function.
$B(t)$	continuous $m \times r$ matrix function.
$C(t)$	$\{\int_0^t \Phi^{-1}(\sigma) f(u(\sigma), \sigma)\, d\sigma \,:\, u(\cdot) \text{ admissible}\}$.
D	$m \times m$ symmetric non-negative definite matrix.
$\overline{D}$	$(m+1) \times (2m+2)$ matrix defined by (3.8.1).
E^n	n-dimensional Euclidean space.
G	$m \times m$ symmetric non-negative definite matrix.
H	convex hull of m points; convex polyhedron.
H_k	convex hull of $y_1(k), \ldots, y_p(k)$, $s(-z_k)$, z_k.
I	identity matrix.
I_k	the set of i, $1 \le i \le p+2$, for which $x^i = 0$.
$I(z)$	$[\min\{\delta\beta(z), 1\}, \ \min\{(2-\delta)\beta(z), 1\}]$.
$J_t(u)$	$\int_0^t [a(\sigma) \cdot x_u(\sigma) + f^0(u(\sigma), \sigma)]\, d\sigma + h^0(x_u(t))$.
K	compact, convex set in E^n
$K(\omega)$	compact, convex sets in E^n continuous on Ω.
$L(x;y)$	$\{z : z = x + \omega(y-x), \ -\infty < \omega < \infty\}$, $x \neq y$; line.
M^i	closed intervals in E^1.
$N(x;\omega)$	$\{z : \lvert z-x \rvert < \omega\}$, $\omega > 0$; open sphere.
$\overline{N}(x;\omega)$	$\{z : \lvert z-x \rvert \le \omega\}$; closed sphere.
O	origin.
$P(y)$	the support hyperplane $\{x : x \cdot y = \eta(y)\}$, $y \neq 0$, of K with outward normal y.

$P(\omega, y)$	the support hyperplane $\{x : x \cdot y = \eta(\omega, y)\}$, $y \neq 0$, of $K(\omega)$ with outward normal y.
$P_H(y)$	the support hyperplane of H with outward normal $y \neq 0$.
$Q(x;y)$	$\{z : z \cdot y = x \cdot y\}$, $y \neq 0$; the hyperplane through x with normal y.
$Q^+(x;y)$	$\{z : z \cdot y < x \cdot y\}$, $y \neq 0$; the open half-space bounded by $Q(x;y)$ with outward normal y.
$Q^-(x;y)$	$\{z : z \cdot y \geq x \cdot y\}$, $y \neq 0$; the closed half-space bounded by $Q(x;y)$ with inward normal y.
$R(t)$	$\{x : x = x_u(t),\ u(\cdot)$ admissible$\}$; reachable set.
S_k	$\{y_1(k), \ldots, y_p(k), s(-z_k)\}$.
$S(y)$	$P(y) \cap K$; the contact set of K for outward normal y.
$S(\omega, y)$	$P(\omega, y) \cap K$; the contact set of $K(\omega)$ for outward normal y.
T	fixed terminal time.
U	compact set in E^r.
$W(t)$	compact, convex, and continuous sets in E^m.
$X(\omega)$	compact, convex, and continuous sets in E^n.
$Y(\omega)$	compact, convex, and continuous sets in E^n.
Y_k	$\{y_1(k), \ldots, y_p(k)\}$.
$a(t)$	continuous function from $[0, T]$ to E^m.
$b(t)$	continuous function from Θ to E^m.
$c_u(t)$	$c_u(t) = \Phi^{-1}(t)\, x_u(t) - x(0)$.
e	the 3-vector (1, 0, 0).
$f(u, t)$	continuous function from $U \times \Theta$ to E^m.
$f^0(u, t)$	continuous function from $U \times \Theta$ to E^1.

g	m-vector in the range of G.
$h^0(x)$	convex function from E^m to E^1.
i	iteration number for GIP.
j	$k + \sum_{q=0}^{i-1} \ell(q)$, $i \neq 0$; k, $i = 0$.
k	iteration number for BIP or IIP.
$\ell(i)$	number of points generated by BIP or IIP (including the initial point) which occur on iteration i of GIP.
p	number of points in Y_k; parameter of IIP.
$s(y)$	function from E^n to K such that $y \cdot s(y) = \max_{z \in K} y \cdot z$, $y \neq 0$; contact function of K.
$s(\omega, y)$	function from E^n to $K(\omega)$ such that $y \cdot s(\omega, y) = \max_{z \in K(\omega)} y \cdot z$, $y \neq 0$; contact function of $K(\omega)$.
$s_X(\omega, y)$	contact function of $X(\omega)$.
t	time.
$u(\cdot)$	r-dimensional control vector function defined on Θ.
$w(T)$	target set consisting of a single point.
$x(t)$	m-dimensional state vector.
x^*	solution of Subproblem 2.
$x_u(t)$	an absolutely continuous solution of (5.1.1) for the admissible control $u(\cdot)$.
y^*	solution of Subproblem 1.
$\tilde{z}(\omega)$	point in $K(\omega)$ satisfying $\lvert \tilde{z}(\omega) \rvert = \min_{z \in K(\omega)} \lvert z \rvert$.
z_i	terminal point generated by BIP or IIP on iteration i of GIP.
$z(j)$	the kth point generated by BIP or IIP on iteration i of GIP.
z^*	solution of BP.

α_k	an element of $I(z_k)$.
$\beta(z)$	$\|z - s(-z)\|^{-2} z \cdot (z - s(-z))$, $z - s(-z) \neq 0$; 0, $z - s(-z) = 0$.
$\gamma(z)$	$\|z\|^{-2} z \cdot s(-z)$, $\|z\| > 0$ and $z \cdot s(-z) > 0$; 0, $z = 0$ or $\|z\| > 0$, $z \cdot s(-z) \leq 0$.
$\gamma(\omega, z)$	$\|z\|^{-2} z \cdot s(\omega, -z)$, $\|z\| > 0$ and $z \cdot s(\omega, -z) > 0$; 0, $z = 0$ or $\|z\| > 0$, $z \cdot s(\omega, -z) \leq 0$.
δ	fixed number in $(0, 1]$ used for $I(z)$; parameter for BIP.
$\bar{\delta}_j$	$\bar{\phi}(z(j) + \omega_i e) - \omega_i$.
$\eta(y)$	$\max_{z \in K} z \cdot y$; support function of K.
$\eta(\omega, y)$	$\max_{z \in K(\omega)} z \cdot y$; support function of $K(\omega)$.
θ	parameter for GIP.
λ_i	principal radii of curvature at z^* for Example 3 of Section 2.5.
$\bar{\lambda}$	$\max_{2 \leq i \leq n} \lambda_i$.
$\mu(z)$	$\|z\|^{-1} z \cdot s(-z)$, $z \neq 0$; 0, $z = 0$.
$\rho(z)$	$\|\tilde{z}(\omega)\|$.
$\bar{\rho}(\underline{x})$	$\|\underline{x} - (\underline{x} \cdot e)e\|$.
$\bar{\rho}_j$	$\bar{\rho}(z(j) + \omega_i e)$.
$\bar{\phi}(\underline{x})$	$\underline{x} \cdot e$.
$\bar{\phi}_j$	$\bar{\phi}(z(j) + \omega_i e)$.
$\phi^i(u, t)$	continuous functions from $U \times \Theta$ to E^1.
$\psi(t)$	m-dimensional adjoint vector.
$\psi(t, y)$	solution of (5.2.5) with boundary condition $\psi(\tau) = y$.
ω	scalar $\in \Omega$.
ω^*	solution of GP.

$\Gamma(z)$	$\|z\|^2 - \|z^*\|^2$, $z \in K$.
∂	denotes "the boundary of"
Δ	denotes "the convex hull of"
$\bar{\Delta}(z;\alpha)$	$\Gamma(z) - \Gamma(z + \alpha(s(-z) - z))$.
Θ	compact interval in E^1.
$\Phi(t)$	solution of $\dot{\Phi} = A(t)\Phi$, $\Phi(0) = I$.
Ω	compact interval in E^1.

Bibliography

A1 Arrow, K.J., Hurwicz, L., and Uzawa, H., "Studies in Linear and Nonlinear Programming," Stanford University Press, Stanford, Calif., (1958).

B1 Babunashvili, T. G., Synthesis of Linear Optimal Systems, (Russian) Doklady Akademii Nauk SSSR, Vol. 155, (1964), pp 295-298, (English translation in SIAM J. on Control, Ser. A, Vol. 2, (1964), pp 261-265).

B2 Barankin, E. W. and Dorfman, R., "Towards Quadratic Programming," Office of Naval Research Logistics Project at Columbia University and University of California, Berkeley, (1955).

B3 Barr, R. O. and Gilbert, E. G., Some Iterative Procedures for Computing Optimal Controls, to appear in "Proceedings of the Third Congress of the International Federation of Automatic Control (IFAC), London, 1966," Butterworth, London.

B4 Beale, E. M. L., On Minimizing a Convex Function Subject to Linear Inequalities, J. of Royal Statistical Society, Ser. B, Vol. 17, (1955), pp 173-184.

B5 Beale, E. M. L., On Quadratic Programming, Naval Research Logistics Quarterly, Vol. 6, (1959), pp 227-243.

B6 Busemann, H., "Convex Surfaces," Interscience Tracts in Pure and Appl. Math. No. 6, Interscience Publishers, Inc., New York, (1958).

B7 Bushaw, D., Optimal Discontinuous Forcing Terms, in "Contributions to the Theory of Nonlinear Oscillations," Princeton University Press, Vol. 4, (1958).

B8 Bellman, R., "Dynamic Programming," Princeton University Press, Princeton, N. J., (1957).

B9 Breakwell, J. V., Optimization of Trajectories, J. SIAM, Vol. 7, No. 2, (1959), pp 215-247.

B10 Breakwell, J. V., Speyer, J. L., and Bryson, A. E., Optimization and Control of Nonlinear Systems Using the Second Variation, SIAM J. on Control, Ser. A, Vol. 1, No. 2, (1963), pp 193-223.

B11 Bryson, A. E., and Denham, W. F., A Steepest-Ascent Method for Solving Optimum Programming Problems, J. Appl. Mech., ASME Trans., Ser. E, Vol. 29, (1962), pp 247-257.

C1 Carathéodory, C., Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen, Rend. Circ. Mat. Palermo, Vol. 32, (1911), pp 193-217.

D1 Dantzig, G. B., Maximization of a Linear Function of Variables Subject to Linear Inequalities. in "Activity Analysis of Production and Allocation," (T. C. Koopmans, editor), Wiley, New York, (1951).

E1 Eaton, J. H., An Iterative Solution to Time-Optimal Control, J. Math. Anal. and Appl., Vol. 5, (1962), pp 329-344; Errata and Addenda, J. Math. Anal. and Appl., Vol. 9, (1964), pp 147-152.

E2 Eggleston, H. G., "Convexity," Cambridge Tracts in Math. and Math. Physics No. 47, Cambridge University Press, (1958).

F1 Fadden, E. J., "Computational Aspects of a Class of Optimal Control Problems," Ph.D. thesis, University of Michigan, 1965.

F2 Fadden, E. J. and Gilbert, E. G., Computational Aspects of the Time-Optimal Control Problem, in "Computing Methods in Optimization Problems," Academic Press, New York, (1964), pp 167-192.

F3 Fancher, P. S., Iterative Computation Procedures for an Optimum Control Problem, IEEE Trans. on Auto. Control, Vol. AC-10, No. 3, (1965), pp 346-348.

F4 Frank, M. and Wolfe P., An Algorithm for Quadratic Programming, Naval Research Logistics Quarterly, Vol. 3, (1956), pp 95-110.

F5 Flugge-Lotz, I., and Marbach, H., The Optimal Control of Some Attitude Control Systems for Different Performance Criteria, J. Basic Engr., ASME Trans., Ser. D, Vol. 85, (1963), pp 165-176.

G1 Gilbert, E. G., An Iterative Procedure for Computing the Minimum of a Quadratic Form on a Convex Set, SIAM J. on Control, Ser. A, Vol. 4, No. 1, (1966), pp 61-80.

G2 Gilbert, E. G., The Application of Hybrid Computers to the Iterative Solution of Optimal Control Problems, in "Computing Methods in Optimization Problems," Academic Press, New York, (1964), pp 261-284.

H1 Hadley, G., "Nonlinear and Dynamic Programming," Addison-Wesley, Reading, Mass., (1964).

H2 Hildreth, C., A Quadratic Programming Procedure, Naval Research Logistics Quarterly, Vol. 4, (1957), pp 79-85.

H3 Ho, Y. C., A Successive Approximation Technique for Optimal Control Systems Subject to Input Saturation, J. Basic Eng., ASME Trans., Ser. D, Vol. 84, No. 1, (1962), pp 33-40.

H4 Houthakker, H. S., The Capacity Method of Quadratic Programming, Econometrica, Vol. 28, (1960), pp 62-87.

H5 Hutcheson, R. D., The Computation of Optimal Controls by an Iterative Procedure, Professional Degree Thesis, University of Michigan, 1965.

K1 Kelley, H. J., Gradient Theory of Optimal Flight Paths, J. American Rocket Society, Vol. 30, (1960), pp 947-953.

K2 Knudsen, H. K., An Iterative Procedure for Computing Optimal Controls, IEEE Trans. on Auto. Control, Vol. AC-9, No. 1, (1964), pp 23-30.

K3 Kuhn, H. W., and Tucker, A. W., Nonlinear Programming, Proceedings Second Berkeley Symposium on Mathematical Statistics and Probability, (1951), pp 481-492.

L1 La Salle, J. P., The Time-Optimal Control Problem, in "Contributions to the Theory of Nonlinear Oscillations," Vol. 5, Princeton University Press, Princeton, N. J., (1960), pp 1-24.

L2 Lemke, C. E., A Method for Solution of Quadratic Programs, Management Science, Vol. 8, (1962), pp 442-453.

M1 Markowitz, H., The Optimization of a Quadratic Function Subject to Linear Constraints, Naval Research Logistics Quarterly, Vol. 3, (1956), pp 111-133.

N1 Neustadt, L. W. and Paiewonsky, B., On Synthesizing Optimal Controls, in "Proceedings of the Second Congress of the International Federation of Automatic Control (IFAC), Basel, 1963", Butterworth, London.

N2 Neustadt, L. W., The Existence of Optimal Controls in the Absence of Convexity Conditions, J. Math. Anal. and Appl., Vol. 7, (1963), pp 110-117.

N3 Neustadt, L. W., Minimum Effort Control Systems, SIAM J. on Control, Ser. A, Vol. 1, (1962), pp 16-31.

N4 Neustadt, L. W., Synthesizing Time Optimal Control Systems, J. Math. Anal. and Appl., Vol. 1, (1960), pp 484-492.

P1 Plant, J. B., and Athans, M., An Iterative Technique for the Computation of Time Optimal Controls, to appear in "Proceedings of the Third Congress of the International Federation of Automatic Control (IFAC), London, 1966," Butterworth, London.

P2 Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., and Mishchenko, E. F., "The Mathematical Theory of Optimal Processes," John Wiley and Sons, Inc., New York, (1962).

V1 Vajda, S., "Mathematical Programming," Addison-Wesley, Reading, Mass., (1961).

V2 Valentine, F. A., "Convex Sets," McGraw-Hill, New York, (1964).

W1 Wolfe, P., The Simplex Method for Quadratic Programming, Econometrica, Vol. 27, (1959), pp 382-398.