SOLVING ASYMMETRIC VARIATIONAL INEQUALITY PROBLEMS AND SYSTEMS OF EQUATIONS WITH GENERALIZED NONLINEAR PROGRAMMING ALGORITHMS

by

JANICE H. HAMMOND

Sc.B. Brown University
(1976)

Submitted to the Department of Mathematics
in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 1984

Signature of Author____________________________
Department of Mathematics
August 31, 1984

Certified by____________________________
Thomas L. Magnanti, Thesis Supervisor

Accepted by____________________________
Nesmith C. Ankeny, Chairman
Departmental Graduate Committee

SOLVING ASYMMETRIC VARIATIONAL INEQUALITY PROBLEMS AND SYSTEMS OF EQUATIONS WITH GENERALIZED NONLINEAR PROGRAMMING ALGORITHMS

by

Janice H. Hammond

Submitted to the Department of Mathematics on August 31, 1984 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

ABSTRACT

The variational inequality problem is a general problem formulation that encompasses a wide range of problems, including optimization problems, complementarity problems, fixed point problems, and network equilibrium problems. In this thesis, we propose and analyze several nonlinear programming-based algorithms for solving variational inequality problems with monotone, asymmetric cost functions. The convergence conditions for most of these algorithms in some way restrict the degree of asymmetry of the Jacobian of the problem map.

Chapter 3 considers a generalization of the steepest descent algorithm for solving unconstrained variational inequality problems. In this chapter, we show that if the problem mapping is affine, then the method converges if M^2, the square of the matrix defining the affine map, is positive definite. We also establish easy to verify conditions on the matrix M that ensure that M^2 is positive definite, and develop a scaling procedure that extends the class of matrices that satisfy the convergence conditions. In addition, we establish a local convergence result for problems defined by nonlinear, uniformly monotone maps, and discuss a class of general descent methods.

Chapter 4 considers several algorithms that generalize first-order approximation methods for solving convex minimization problems. We device a "contracting ellipsoid" method that solves a variational inequality problem by solving a sequence of quadratic programming problems. We prove convergence of this algorithm for constrained problems defined by uniformly monotone maps, and geometrically interpret the algorithm in terms of a sequence of ellipsoidal level sets. The results of the chapter also show that the subgradient algorithm for solving convex, non-differentiable minimization problems can be used to solve a max-min problem that is equivalent to the variational inequality problem. Finally, the chapter discusses a generalization of the Frank-Wolfe method for solving variational inequality problems, and shows that a variant of the Frank-Wolfe method will solve a certain class of variational inequality problems.

Thesis Supervisor: Thomas L. Magnanti
Title: Professor of Operations Research and Management

ACKNOWLEDGEMENTS

I would like to thank the members of my thesis committee, Tom Magnanti, Dimitri Bertsekas, and Rob Freund, for their advice and contributions to this thesis. I would especially like to acknowledge the considerable contribution by Tom Magnanti to the content and exposition of this work. I have been very fortunate to have had Tom as my thesis advisor. The many hours we have spent discussing this work have been both intellectually stimulating and fun. I have benefited greatly from these discussions and from Tom's thoughtful advice. I also deeply appreciate his kindness and the encouragement he has given me during the past four years.

I would like to thank Leslie Regan and Kate Sumera for their excellent and efficient typing of this thesis, and Vijay Singhal for his generous assistance in proofreading it. I would also like to thank Phyllis Ruby and Marcia Ryder for their friendship and administrative help, and Al Drake for his encouragement and support that have been so helpful in the development of my teaching skills.

My decision to pursue a Ph.D. was partly inspired and sustained by my friendships with a few delightful and goodhearted mathematicians, especially Carl Shapiro, Ben Thompson, Jeff Hoffstein and Sheldon Axler. Several very special friends have greatly enriched my life during the past few years: the members of the MIT Women's Rugby Club, especially Connie Cepko and Martha Nelson; Karen Johnson; and Scott Shepard. I would like to thank my family for their love and support: my mother and father; my sister, Carolyn; and my "honorary aunts," Edith Tremel and Madeline Taylor. I would especially like to thank my parents for their support, for providing me with so many opportunities, and for never once suggesting that math was an inappropriate pursuit for their daughter. And finally, to Cliff Tabin, who has done so very much for me: thank you.

This research was supported in part by the National Science Foundation's Program in Systems Theory and Operations Research under Grants ECS-792-6225 and ECS-831-6624, and in part by a fellowship, endowed by Louis Seley, from the Transportaion Association of America.

TABLE OF CONTENTS

Page

TITLE PAGE.. 1

ABSTRACT.. 2

ACKNOWLEDGEMENTS.. 3

TABLE OF CONTENTS... 4

LIST OF FIGURES... 6

CHAPTER 1 - INTRODUCTION...................................... 8

1.1 Historical Development of the Problem..................... 8
1.2 Statement of the Problem.................................. 10
1.3 Motivation for Adapting Nonlinear Programming Algorithms to Solve Variational Inequalities: The Role of Symmetry of the Jacobian of f.................................. 12
1.4 Applications.. 14
1.5 Notation and Definitions.................................. 20
1.6 Outline of the Thesis..................................... 24

CHAPTER 2 - LITERATURE SURVEY................................. 26

2.1 Projection Methods.. 27
2.2 Nonlinear Programming-Based Algorithms.................... 34
2.3 Simplicial Decomposition Algorithms....................... 49
2.4 Algorithms Based on Reformulating the Variational Inequality Problem as a Max-Min Problem.................. 55

CHAPTER 3 - THE GENERALIZED STEEPEST DESCENT METHOD FOR UNCONSTRAINED VARIATIONAL INEQUALITY PROBLEMS.............. 69

3.1 Introduction.. 69
3.2 The Generalized Steepest Descent Algorithm for Unconstrained Problems with Affine Maps....................... 75
3.3 The Generalized Steepest Descent Algorithm for Unconstrained Problems with Nonlinear Maps.................... 95
3.4 Scaling the Mapping of an Unconstrained Variational Inequality Problem....................................... 106
3.5 Generalized Descent Algorithms............................ 113
3.6 Concluding Remarks.. 119

CHAPTER 4 - GENERALIZATIONS OF FIRST-ORDER APPROXIMATION METHODS.. 120

4.1 A Contracting Ellipsoid Algorithm and Its Interpretation.. 121
4.2 Subgradient Algorithms.................................... 153
4.3 The Frank-Wolfe Algorithm................................. 159
4.4 Concluding Remarks.. 175

Page

CHAPTER 5 - CONCLUSION.. 177

REFERENCES... 183

LIST OF FIGURES

FIGURE NUMBER	TITLE	PAGE
2.1	Variational Inequality Problem as a Fixed Point Problem	27
2.2	$H_k(x)$ is a Piecewise Linear Approximation to $H(x)$	59
3.1	The Steepest Descent Iterates Need Not Converge If M is Asymmetric	73
3.2	A Comparison of Movement Directions for Different Values of p	74
3.3	M^2 is Positive Definite If and Only If $(M^Tx)^TMx>0$ for Every Nonzero x	82
4.1	The Contracting Ellipsoid Method Solves An Unconstrained Affine Variational Inequality Problem by Generating a Sequence of Ellipsoids that Contract to the Solution x^*	126
4.2	The Approximate Map $g(\bar{x},x^k)$ Points Away from x^* if $\|\bar{x}-x^*\|_S > c\|x^k - x^*\|_S$	133
4.3	The Contracting Ellipsoid Method Solves a Constrained Affine Variational Inequality Problem	136
4.4	Relationship Between the Ellipsoid Level Sets and the Steepest Descent Method	145
4.5	The Contracting Ellipsoid Iterates Move Halfway to the Solution if M is Symmetric	146
4.6	The Generalized Frank-Wolfe Algorithm Cycles	163
4.7	The Generalized Frank-Wolfe Method Cycles on a Traffic Equilibrium Problem	165
4.8	Frank-Wolfe Iterates Need Not Contract to the Solution	167
4.9	$f(x) = Mx$ is Normal to the Tangent Plane to the Level Set if and Only if M is Symmetric	168
5.1	The Direction $-f(x^k)$ is not a Good Movement Direction if the Jacobian of f(x) is Very Asymmetric	179

FIGURE NUMBER	TITLE	PAGE
5.2	A Full Steplength Pulls the Iterate Further From the Solution When the Map is Very Asymmetric	180

CHAPTER 1

INTRODUCTION

The variational inequality problem is a general problem formulation that encompasses a wide range of problems, including, among others, optimization problems, complementarity problems, fixed point problems, and network equilibrium problems.

In this thesis, we propose and analyze several nonlinear programming-based algorithms for solving variational inequality problems with monotone, asymmetric cost functions. This introductory chapter first provides a brief history of the development of the variational inequality problem. Then, after stating the general variational inequality problem, we define a number of conditions that are often imposed on the problem data, and discuss the existence and uniqueness of solutions to the problem. The following section motivates the use of nonlinear programming-based algorithms as solution procedures by establishing conditions under which the problem is equivalent to a convex minimization problem. We then discuss several broad categories of problem types that can be formulated as variational inequalities, and describe in detail the reformulation of the topical traffic equilibrium problem as a variational inequality problem. Finally, we provide some background definitions and list the notational conventions to be used in the thesis. The chapter ends with a brief outline of the thesis.

1.1 Historical Development of the Problem

The theory and methodology of the variational inequality problem originated primarily from studies of certain classes of partial differential equations. In particular, much of the early work on the problem focused on the formulation and solution of free boundary value problems. In these settings, the problem is usually formulated over an infinite dimensional function space; in contrast, this work will discuss problems formulated over finite dimensional spaces. In this setting, the variational inequality problem is of particular interest to mathematical programmers because it includes as special cases virtually all of the classical problems of mathematical programming: convex programming problems, network equilibrium problems, linear and nonlinear complementarity problems, fixed point problems, and minimax problems.

Mathematical programmers' recent interest in the variational inequality problem stems primarily from the recognition that the equilibrium conditions for network equilibrium problems, such as the traffic equilibrium problem and the spatially separated market equilibrium problem, can be formulated in a natural way as a variational inequality problem. For example, in 1980, Dafermos [1980] recognized that the traffic equilibrium conditions formulated by Smith [1979] defined a variational inequality problem.

In the past few years, motivated by the desire to find methods to solve such equilibrium problems, a number of researchers have developed algorithms to solve variational inequality problems. The algorithms tend to fall into four general categories: projection algorithms, nonlinear programming algorithms adapted to the variational inequality problem, simplicial decomposition algorithms, and algorithms to solve an equivalent max-min problem. Chapter 2 of this work surveys some of the recent research on the development of these algorithms.

1.2 Statement of the Problem

The variational inequality problem, VI(f,C), over a ground set $C \subseteq R^n$ is a specially structured system of nonlinear inequalities; namely,

Find $x^* \in C$ satisfying $(x - x^*)^T f(x^*) \geq 0$ for every x in C, (VI(f,C))

where $C \subseteq R^n$ is generally assumed to be convex and compact (or, in some cases, convex and closed), and $f: C \rightarrow R^n$ is generally assumed to be

(1) either continuous, hemicontinuous (i.e., continuous along straight lines), or continuously differentiable; and

(2) either

(a) monotone on C: $(x - y)^T (f(x) - f(y)) \geq 0$ for every $x \in C$, $y \in C$;

(b) strictly monotone on C: $(x - y)^T (f(x) - f(y)) > 0$ for every $x \in C$, $y \in C$ with $x \neq y$;

(c) uniformly (or strongly) monotone on C: for some scalar $k > 0$, $(x - y)^T (f(x) - f(y)) \geq k||x - y||^2$ for every $x \in C$, $y \in C$, where $||\cdot||$ denotes the Euclidean norm; or

(d) monotone on C with strict monotonicity at the solution x^*.

We will refer to any $x^* \in C$ satisfying $(x - x^*)^T f(x^*) \geq 0$ for all $x \in C$ as a solution to VI(f,C).

More generally, VI(f,C) can be formulated over a real Hilbert space H with inner product $(\cdot,\cdot)$. Let C be a closed convex subset of H and f be a mapping from H into its dual H'. In this case, the problem becomes,

Find $x^* \in C$ satisfying $(x - x^*, f(x^*)) \geq 0$ for every $x \in C$.

The following results specify conditions ensuring that VI(f,C) has a solution. (Note that if we eliminate the assumption that C is bounded, ensuring the existence of a solution requires that either a uniform monotonicity condition or a coercivity condition be imposed on f.)

Theorem 1.1 (Existence) (Kinderlehrer and Stampacchia [1980])

If $C \subseteq R^n$ is compact and convex, and $f:C \to R^n$ is continuous, then the variational inequality problem VI(f,C) has a solution.

Theorem 1.2 (Existence) (Kinderlehrer and Stampacchia [1980])

Suppose that $C \subseteq R^n$ is closed and convex, and $f:C \to R^n$ is continuous and satisfies the following coercivity condition: there exists an $x^0 \in C$ such that

$$\lim_{\substack{||x|| \to \infty \\ x \in C}} \frac{(x-x^0)^T[f(x)-f(x^0)]}{||x-x^0||} = +\infty.$$

Then, the variational inequality problem VI(f,C) has a solution.

Theorem 1.3 (Existence) (Auslender [1976])

Suppose that $C \subseteq R^n$ is closed and convex, and $f:C \to R^n$ is monotone, hemicontinuous, and satisfies the following coercivity condition on C: there exists an $x^0 \in C$ and a scalar $k > 0$ such that

$$\text{if } x \in C \text{ and } ||x|| > k, \text{ then } (x-x^0)^T f(x) > 0.$$

Then, the variational inequality problem VI(f,C) has a solution.

Theorem 1.4 (Existence) (Auslender [1976])

If $C \subseteq R^n$ is closed and convex, and f is uniformly monotone and hemicontinuous on C, then the variational inequality problem VI(f,C) has a solution.

The following theorem specifies conditions on the mapping f that ensure that the solution x^* is unique.

Theorem 1.5 (Uniqueness)

Let x^* be a solution to VI(f,C). If f is monotone on C and strictly monotone at x^*, then the solution x^* is unique.

Clearly, then, if the mapping f is either strictly or uniformly monotone on C, the solution will be unique.

1.3 Motivation for Adapting Nonlinear Programming Algorithms to Solve Variational Inequalities: The Role of Symmetry of the Jacobian of f

The mapping $f:C \subseteq R^n \to R^n$ is a gradient mapping on C (or is integrable on C) if there exists a Gateaux differentiable functional $F:C \subseteq R^n \to R^1$ such that $[\nabla F(x)]^T = f(x)$ for every x in C. Gradient mappings can be characterized as follows:

Symmetry Principle: (Ortega and Rheinboldt [1970])

Let $f:D \subseteq R^n \to R^n$ be continuously differentiable on an open convex subset $D' \subseteq D$. Then f is a gradient mapping on D' if and only if $\nabla f(x)$ is symmetric for every x in D'.

If f is the gradient of a convex, continuously differentiable functional $F:C \subseteq R^n \to R^1$, then x^* solves the variational inequality problem VI(f,C) precisely when x^* minimizes the functional F over C. To demonstrate this result, we first note that since F is convex and Gateaux differentiable on C, then $F(x) - F(y) \geq \nabla F(y)(x-y) = f^T(y)(x-y) = (x-y)^T f(y)$ for every x and y in C. Thus, if $x^* \varepsilon C$ satisfies $(x-x^*)^T f(x^*) \geq 0$ for every x in C, then $F(x) - F(x^*) \geq 0$ for every x in C, and, hence, x^* minimizes F over C. Conversely, if x^* minimizes F over C, then the directional derivative of F at x^* in all feasible directions from x^* must be nonnegative; i.e., for every $x \varepsilon C$, x^* must satisfy the gradient condition $\nabla F(x^*)(x-x^*) = f^T(x^*)(x-x^*) = (x-x^*)^T f(x^*) \geq 0$.

If $f: C \subseteq R^n \rightarrow R^n$ is the gradient of a Gateaux differentiable functional $F: C \subseteq R^n \rightarrow R^1$, then the following statements specify the relationship between the convexity of F and the monotonicity of f:

(i) f is monotone on C if and only if F is convex on C;

(ii) f is strictly monotone on C if and only if F is strictly convex on C; and

(iii) f is uniformly monotone on C if and only if F is uniformly convex on C.

These statements and the symmetry condition stated earlier show that the variational inequality problem is equivalent to a convex minimization problem with a (strictly, uniformly) convex objective function whenever f is a (strictly, uniformly) monotone, continuously differentiable mapping with a symmetric Jacobian on C. Whenever f satisfies these properties, we can solve the variational inequality problem using any algorithm that will solve the equivalent convex minimization problem, either by determining the function F(x) for which $f(x) = \nabla F(x)$ and using the algorithm to solve the equivalent optimization problem, or, in some cases, by adapting the algorithm to the variational inequality formulation and solving the problem directly.

This thesis studies the following question: When can nonlinear programming algorithms be adapted to solve the variational inequality problem when f is monotone, but not necessarily a gradient mapping? Research by Ahn [1979], Dafermos [1982a], [1983], and Pang and Chan [1982] show that certain nonlinear programming algorithms adapted to solve the variational inequality problem will converge when f is not necessarily a gradient mapping as long as f satisfies certain diagonal dominance conditions. The literature survey in Chapter 2 reviews these methods and other algorithms for solving variational inequality problems.

1.4 Applications

As we have mentioned earlier, the variational inequality formulation encompasses a number of problem types. For this reason, a wide range of problem applications can be cast as variational inequality problems. In this section, we discuss the relationship between the variational inequality formulation and a number of classical mathematical programming problems. We first exhibit the correspondences between variational inequality problems and complementarity problems, fixed point problems, nonlinear equations, and minimization problems. Because the reformulation of the traffic equilibrium problem as a variational inequality problem has stimulated much of the recent research activity on variational inequality problems, we then focus our discussion of applications on this reformulation.

1.4.1 Relationship to other Problems

Each of the following well-known problems is closely related to the variational inequality problem. The results below demonstrate these relationships. (Proofs of these results appear in Cottle [1974], Kinderlehrer and Stampacchia [1980] and Lemke [1980].)

Complementarity Problems:

Let R^n_+ denote the nonnegative orthant in R^n, and let $f:R^n \to R^n$. The nonlinear complementarity problem over R^n_+ is a system of equalities and inequalities stated as,

$$\text{Find } x^* \geq 0 \text{ such that } f(x^*) \geq 0 \text{ and } (x^*)^T f(x^*) = 0. \qquad (CP(f))$$

The nonlinear complementarity problem is a linear complementarity problem whenever the mapping f is affine, i.e., whenever $f(x) = Mx - b$, where M is an $n \times n$ matrix and b an $n \times 1$ vector.

Theorem 1.6

$VI(f,R^n_+)$ and $CP(f)$ have precisely the same solutions, if any.

Fixed Point Problems:

Let $C \subseteq R^n$ be a closed convex subset of R^n, and let $g:C \to C$. The fixed point problem is a specialized system of equations; namely,

$$\text{Find } x^* \in C \text{ such that } x^* = g(x^*). \qquad (FP(g,C))$$

Theorem 1.7

Let $C \subseteq R^n$ be closed and convex.

(i) Let $g:C \to C$, and let $f(x) = x - g(x)$ for every $x \in C$. Then $VI(f,C)$ and $FP(g,C)$ have precisely the same solutions, if any.

(ii) Let $f:C \to R^n$, and let $g(x) = P_C[x - f(x)]$ for every $x \in R^n$, where P_C is the projection operator onto the set C. Then $VI(f,C)$ and $FP(g,C)$ have precisely the same solutions, if any.

Equations:

Let $C \subseteq R^n$ and $h:C \to R^n$. An equation simply seeks a zero of the mapping h over C, i.e.,

$$\text{Find } x^* \in C \text{ such that } h(x^*) = 0. \qquad (EQ(h,C))$$

Theorem 1.8

Let C be closed and convex, let $f:C \to R^n$, and let $h(x) = x - P_C[x - f(x)]$ for every $x \in C$. Then VI(f,C) and EQ(h,C) have precisely the same solutions, if any. (Note that if $h(x) = x - g(x)$, then the solutions of EQ(h,C) and FP(g,C) are the same.)

Minimization Problems:

Let $C \subseteq R^n$ be closed and convex, and $F:C \to R^1$. We seek a minimum of F over C, i.e.,

$$\text{Find } x^* \in C \text{ such that } F(x^*) \leq F(x) \text{ for every } x \in C. \qquad (MIN(F,C))$$

Theorem 1.9

Let $C \subseteq R^n$ be closed and convex, let $F:C \to R^1$ be continuously differentiable, and let $f(x) = \nabla F(x)$.

(i) If x^* is a solution to MIN(F,C) then x^* is a solution to VI(f,C).

(ii) If F is convex, then MIN(F,C) and VI(f,C) have precisely the same solutions, if any.

1.4.2 Formulation of the Traffic Equilibrium Problem as a Variational Inequality Problem

Many network equilibrium problems can be formulated as variational inequality problems. In this subsection, we describe such a formulation for one network equilibrium problem, the traffic equilibrium problem.

(Another important setting for the network equilibrium problem is the spatially separated economic market equilibrium problem; see, for example, Florian and Los [1981], Samuelson [1952], and Takayama and Judge [1971].)

For simplicity, we consider a version of the traffic equilibrium problem with fixed demand, which is usually referred to as the traffic assignment problem. The following analysis can, however, be extended to the more general problem with elastic demand (see Dafermos [1982b]).

Let $G = [N,A]$ be a network consisting of a set N of nodes and a set A of directed arcs. Let W be a set of origin-destination node pairs. For each $w \in W$, let P_w be the set of directed paths joining the OD (origin-destination) pair w. We assume a fixed demand, d_w, for travel from the origin node to the destination node of OD pair w. Let x be a vector of path flows:

$$x = (x_p)_{p \in P_w, w \in W}.$$

Let C be the set of all feasible path flow vectors:

$$C = \{x: \sum_{p \in P_w} x_p = d_w,\ x_p \geq 0 \text{ for every } p \in P_w \text{ and } w \in W\}.$$

Finally, we let $t_p(x)$ represent the marginal cost (i.e., travel time) of a unit of flow on path p as a function of the flow x on the network.

In this setting, we will define a user equilibrium flow pattern to be a flow having the property that no user has an incentive to unilaterally change his or her route. (Wardrop [1952] first formalized this behavioral assumption.) If we assume that a user's sole criterion for route selection is travel time, then this equilibrium flow must satisfy the property that for a given OD pair w, if a path $p \in P_w$ is used (i.e., if the flow x_p is positive), then the travel time $t_p(x)$ on that path must be minimal among travel times on all paths joining OD pair w. This equilibrium principle can be stated mathematically as follows: x^* is an equilibrium flow if for each $w \in W$ and each $\bar{p} \in P_w$, it satisfies

$$\text{if } x^*_{\bar{p}} > 0, \text{ then } t_{\bar{p}}(x^*) = \text{Min}\{t_p(x^*) : p \in P_w\}. \tag{1.1}$$

This condition states that for a given OD pair, the travel times on all <u>used</u> paths connecting this OD pair are equal and that this travel time does not exceed the travel time that would be incurred if one vehicle were to travel on any <u>unused</u> path connecting this OD pair.

Following Smith [1979], we now show that these equilibrium conditions can be reformulated as the variational inequality problem:

Find $x^* \in C$ such that $(x - x^*)^T t(x^*) \geq 0$ for every $x \in C$,

or, equivalently,

$$(x^*)^T t(x^*) \leq x^T t(x^*) \text{ for every } x \in C. \tag{1.2}$$

We may interpret the equation (1.2) as follows: with costs on the network fixed at $t(x^*)$, the total cost $(x^*)^T t(x^*)$ for users in flow pattern x^* is less than or equal to the total cost $x^T t(x^*)$ for users in any other feasible flow pattern x. To show that the equilibrium conditions 1.1 are equivalent to the conditions 1.2, first note that if x^* satisfies the conditions 1.1, then for each OD pair w, the flow x^* uses only shortest paths (with respect to costs $t(x^*)$). If this is true, we cannot reduce the total cost (again with respect ot costs $t(x^*)$) of travel on the system by changing the flow pattern to any other feasible flow. Consequently, x^* satisfies 1.2. Conversely, suppose that x^* does not satisfy the equilibrium conditions 1.1. Then there exists an OD pair w and a path $p \in P_w$ such that $x_p^* > 0$ and $t_p(x^*) > t_{p'}(x^*)$ for some $p' \in P_w$. So with travel costs fixed at $t(x^*)$, we could reduce the total travel time of the users by diverting some of the flow from path p to path p'. The resulting flow pattern x' would have a lower total travel time (with respect to costs $t(x^*)$) than the flow pattern x^*, i.e., $(x^*)^T t(x^*) > (x')^T t(x^*)$. Thus, if x^* does not satisfy 1.1, then it does not satisfy 1.2, and, therefore, the two formulations are equivalent. As a result, $x^* \in C$ is a user-equilibrium flow pattern if and only if $(x - x^*)^T t(x^*) \geq 0$ for every $x \in C$.

Solution methods for the traffic equilibrium problem have evolved through three "generations." The first methods for finding equilibrium flow patterns attempted to incrementally load the network by adding flow to paths that were shortest with costs defined with respect to the flow already loaded. In doing so, these heuristic procedures attempted to use paths for a given OD pair that were approximately equal in travel time and that were lower in travel time than all unused paths for that OD pair. The underlying models were "separable": they assumed that the travel time on any given arc in the arc set was a function of the flow on that arc only--not on the flow on the entire network.

The discovery by Beckman, Winston, and McGuire [1956] that the equilibrium conditions for the separable model could be interpreted as the optimality conditions for a convex minimization problem attracted the interest of operations researchers and spurred the development of a number of optimization-based algorithms. See, for example, Dafermos [1971], [1972], LeBlanc et al. [1975], and Nguyen [1974]. (The interpretation of the equilibrium conditions as the optimality conditions for a convex minimization problem is also valid for the nonseparable model if the Jacobian of the travel cost function is symmetric over C.)

The more recent discovery that the problem can be formulated as a variational inequality problem has once again motivated researchers to study the problem and develop new algorithms. The variational inequality formulation is an important extension of the previous formulations because it allows considerably more flexibility in the problem formulation than the separable model. Both the separable and nonseparable models allow for "elastic" demand: the demand for flow between OD pair w is a function of the shortest path times on that path. The variational inequality formulation, however, allows a number of additional modelling extensions, including

(1) asymmetric travel costs;

(2) multiple modes of transit;

(3) multiple user classes;

(4) link interactions; and

(5) destination choice, trip generation, and other complex demand models.

Aashtiani and Magnanti [1980] and Magnanti [1982] specify how to incorporate these modelling features into the general model.

1.5 Notation and Definitions

In this section we briefly outline the notational conventions and terminology to be used in this work. Other definitions and notations will be introduced in the text as needed.

Let M be a real nxn matrix. In general, we define the definiteness of M without regard to symmetry:

(i) M is positive definite if and only if $x^T Mx > 0$ for every nonzero $x \in R^n$; and

(ii) M is positive semidefinite if and only if $x^T Mx \geq 0$ for every $x \in R^n$.

If we let $\hat{M}$ denote the symmetric part of the matrix M; i.e.,

$$\hat{M} := M + M^T,$$

where T denotes transposition and := denotes definition, then, since $x^T Mx = x^T \hat{M} x$ for every $x \in R^n$,

(i) M is positive definite if and only if $\hat{M}$ is positive definite; and

(ii) M is positive semidefinite if and only if $\hat{M}$ is positive semidefinite.

An nxn positive definite symmetric matrix defines an inner product on R^n:

$$(x,y)_G := x^T Gy.$$

The inner product defined by G induces a norm on R^n:

$$||x||_G := (x,x)_G^{\frac{1}{2}} = (x^T Gx)^{\frac{1}{2}},$$

which, in turn, induces a norm on the nxn matrix A:

$$\|A\|_G := \sup_{\|x\|_G=1} \|Ax\|_G.$$

(By writing $\|A\|_G$, it is implicit that A has the same dimensions as G.) The G norm also induces a projection operator on a subset C of R^n:

$$P_C^G(z) := \underset{x \in C}{\operatorname{argmin}} \|x - z\|_G,$$

where

$$\underset{x \in C}{\operatorname{argmin}} f(x) := \{\bar{x}: f(\bar{x}) = \min_{x \in C} f(x)\}.$$

If C is convex and closed, then $P_C^G(z)$ is the unique point in C that is closest, with respect to the G norm, to z.

If G is the identity matrix I, then the above definitions become the Euclidean inner product, $(\cdot,\cdot)$; norm, $\|\cdot\|_2$; and projection operator, $P_C(\cdot)$.

Let $G^{\frac{1}{2}}$ be any matrix satisfying $(G^{\frac{1}{2}})^T G^{\frac{1}{2}} = G$. Then, the following equivalences follow from the definitions of $\|\cdot\|_G$. For any $x \in R^n$ and any nxn matrix A,

$$\|x\|_G = \|G^{\frac{1}{2}} x\|_2;$$

$$\|A\|_G = \|G^{\frac{1}{2}} A (G^{\frac{1}{2}})^{-1}\|_2; \text{ and}$$

$$\|A\|_2 = \|(G^{\frac{1}{2}})^{-1} A G^{\frac{1}{2}}\|_G.$$

A positive definite symmetric matrix G always has a square root matrix $G^{\frac{1}{2}}$ satisfying $(G^{\frac{1}{2}})^T G^{\frac{1}{2}} = G$. The square root of G is not unique. For example, let $G^{\frac{1}{2}} = U^T D^{\frac{1}{2}} U$, where D is a diagonal matrix whose diagonal entries are the eigenvalues of G and U is an orthogonal matrix whose columns are the corresponding eigenvectors of G. Then, $(G^{\frac{1}{2}})^T G^{\frac{1}{2}} = U^T DU = G$. A square root matrix defined in this way is symmetric. In general, however, the square root of G can be asymmetric: for example, the Cholesky decomposition of G is generally not symmetric.

Let $\lambda(M)$ denote the set of eigenvalues of the nxn real matrix M. If M is positive definite, then

$$\lambda_{min}(\hat{M})\ x^T x \leq x^T M x \leq \lambda_{max}(\hat{M})\ x^T x \qquad \text{for every } x \in R^n,$$

where

$\lambda_{min}(\hat{M}) :=$ minimum eigenvalue of $\hat{M}$, and

$\lambda_{max}(\hat{M}) :=$ maximum eigenvalues of $\hat{M}$.

Because $\hat{M}$ is symmetric and positive definite, all of the eigenvalues of $\hat{M}$ are real and positive, and, hence, the minimum and maximum eigenvalues are well-defined.

If A is an nxn nonsingular matrix, then $(A^{-1})^T = (A^T)^{-1}$, so the definition

$$A^{-T} := (A^{-1})^T = (A^T)^{-1}$$

is well-defined.

Finally, if $f:R^n \to R^n$ is Gauteaux differentiable, we let $\nabla f(x)$ denote the nxn Jacobian matrix of f at x. The $(i,j)^{th}$ entry of $\nabla f(x)$ is given by

$$[\nabla f(x)]_{ij} = \frac{\partial}{\partial x_j} f_i(x).$$

If $F:R^n \to R^1$ is Gauteaux differentiable, the gradient of F at x is the transpose of the column vector $\nabla F(x)$ with components defined by

$$[\nabla F(x)]_j = \frac{\partial}{\partial x_j} F(x), \qquad j = 1,2,\ldots,n.$$

1.6 Outline of the Thesis

In this section, we briefly review the contents of the thesis.

Chapter 2 contains a unified summary of recent algorithmic research on variational inequality problems. The review focuses on algorithms designed to solve variational inequality problems that arise in network equilibrium settings.

Chapter 3 considers a generalization of the steepest descent algorithm for solving unconstrained variational inequality problems. In this chapter, we show that if the problem mapping is affine, then the method converges if M^2, the square of the matrix defining the affine map, is positive definite. We also establish easy to verify conditions on the matrix M that ensure that M^2 is positive definite, and develop a scaling procedure that extends the class of matrices that satisfy the convergence conditions. In addition, we establish a local convergence result for problems defined by nonlinear, uniformly monotone maps, and discuss a class of general descent methods.

Chapter 4 considers several algorithms that generalize first-order approximation methods for solving convex minimization problems. We devise a "contracting ellipsoid" method that solves a variational inequality problem by solving a sequence of quadratic programming problems. We prove convergence of this algorithm for constrained problems defined by uniformly monotone maps, and geometrically interpret the algorithm in terms of a sequence of ellipsoidal level sets. The results of the chapter also show that the subgradient algorithm for solving convex, nondifferentiable minimization problems can be used to solve a max-min problem that is equivalent to the variational inequality problem. Finally, the chapter discusses a generalization of the Frank-Wolfe method for solving variational inequality problems, and shows that a variant of the Frank-Wolfe method will solve a certain class of variational inequality problems.

Chapter 5 contains general conclusions that are drawn from the results of the thesis. We describe geometrically the difficulties that arise when nonlinear programming algorithms are adapted to solve variational inequality problems, and show how we can avoid these problems by restricting the degree of asymmetry of the Jacobian of the underlying problem map. Chapter 5 also suggests directions for future research.

CHAPTER 2

LITERATURE SURVEY

This chapter reviews recent algorithmic research on variational inequality problems. We divide the algorithms into four general categories: (i) projection algorithms, (ii) nonlinear programming-based algorithms, (iii) simplicial decomposition algorithms, and (iv) algorithms solving a max-min problem that is equivalent to the variational inequality problem. The purpose of this review is to present a unified summary of this algorithmic research that will acquaint the reader with existing algorithms for solving variational inequality problems, and, in doing so, will provide a contextual framework in which to view the results of this thesis.

In order to motivate each type of algorithm, we begin each of the following sections with a general overview of the class of algorithms described in the papers reviewed in that section. We then discuss the specifics of the algorithms described in each paper: the statement of the algorithm, the assumptions under which the algorithm converges, the problem settings for which the algorithm is well-suited, and the relationship between the algorithm and others.

Because the literature in the field is extensive, we focus our discussion on algorithms designed to solve variational inequality problems that arise in network equilibrium settings. Earlier research on variational inequalities focused on the solution of partial differential equations. (See, for example, Hartman and Stampacchia [1966], Lions and Stampacchia [1967], Browder [1966], and Sibony [1970].) For earlier work on the traffic equilibrium problem, see Beckman et al. [1956], Dafermos [1971], LeBlanc et al. [1975], Leventhal et al. [1973], Nguyen [1974], Florian [1976], and Magnanti [1982].

2.1 Projection Methods

The variational inequality problem VI(f,C) is equivalent to the fixed point problem

$$\text{Find } x^* \in C \text{ satisfying } x^* = P_C^G[x^* - G^{-1}f(x^*)],$$

where G is any given symmetric positive definite matrix and P_C^G is the projection operator onto the set C with respect to the G-norm. To show that these problems are equivalent, we first recall that z is the projection of y onto the closed convex set C with respect to the G-norm if and only if $z \in C$ and $(x-z,\ z-y)_G = (x-z)^T G(z-y) \geq 0$ for every $x \in C$. (See, for example, Kinderlehrer and Stampacchia [1980].) Consequently, $x^* = P_C^G[x^* - G^{-1}f(x^*)]$ if and only if $x^* \in C$ and $(x-x^*,\ x^* - [x^* - G^{-1}f(x^*)])_G = (x-x^*,\ G^{-1}f(x^*))_G = (x-x^*)^T f(x^*) \geq 0$ for every $x \in C$. Figure 2.1 illustrates this equivalence when $G = I$.

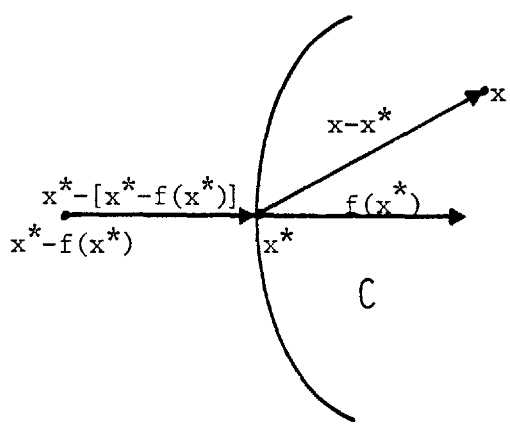


Figure 2.1

Variational Inequality Problem as a Fixed Point Problem

The reformulation of VI(f,C) as finding the fixed point of a projection mapping motivates the use of projection algorithms to solve VI(f,C). The general framework for these algorithms can be stated as follows:

<u>Projection Algorithm</u>

Step 0: Select $x^o \varepsilon$ C. Set k = 0.

Step 1: Given $x^k \varepsilon$ C, let

$$x^{k+1} = P_C^G[x^k - wG^{-1}f(x^k)],$$

where w may be considered a "steplength."

Return to Step 1 with k = k+1. ∎

This projection algorithm will find the solution x^* to VI(f,C) if: (Sibony [1970])

(1) f is Lipschitz continuous and uniformly monotone;

(2) C is closed and convex; and

(3) the stepsize w is sufficiently small.

Since G is a positive definite symmetric matrix, the projection in step 1, stated as minimizing the squared G-norm of $x - [x^k - wG^{-1}f(x^k)]$ over C, reduces, after dividing by 2w and deleting the constant term, to the convex minimization problem:

$$\text{Min } (x-x^k)^T f(x^k) + (1/(2w))(x-x^k)^T G(x-x^k).$$

In general, the rate of convergence of these projection algorithms is linear.

The first two papers in this section focus on the use of projection methods for solving variational inequality problems in the traffic equilibrium setting. The third paper discusses a method to adapt the projection method to solve variational inequality problems with monotone, but not necessarily uniformly monotone, mappings.

2.1.1 Dafermos: "Traffic Equilibrium and Variational Inequalities."

This paper suggests solving the variational inequality problem VI(f,C) by solving a sequence of linear variational inequality problems that can be reformulated as equivalent optimization problems. Let $f^k(x) = G(x-x^k) + wf(x^k)$, where G is a fixed positive definite symmetric matrix and the "stepsize" w is a positive scalar. The mapping $f^k(x)$ is a linear approximation to f(x) about the point x^k.

Algorithm

Step 0: Select $x^o \in C$. Set $k = 0$.

Step 1: Given $x^k \in C$, let x^{k+1} solve the variational inequality problem $VI(f^k,C)$.

If x^k is a solution to $VI(f^k,C)$, stop: x^k solves VI(f,C).

Otherwise, return to Step 1 with $k = k+1$. ∎

(To see that x^k solves VI(f,C) if x^k solves $VI(f^k,C)$, note that $(x-x^k)^T f^k(x^k) = (x-x^k)^T[G(x^k-x^k) + wf(x^k)] = w(x-x^k)^T f(x^k)$.)

Although not stated as such, this algorithm is a projection algorithm. Because $\nabla f^k(x) = G$ is a positive definite symmetric matrix, the k^{th} subproblem $VI(f^k,C)$ is equivalent to the convex minimization problem

$$\operatorname*{Min}_{x\varepsilon C} \ (1/2)x^T Gx - x^T Gx^k + wx^T f(x^k),$$

which has the same solution as the convex minimization problem

$$\min_{x \in C} (1/(2w))(x-x^k)^T G(x-x^k) + (x-x^k)^T f(x^k),$$

since the objective function of the first problem is the objective function of the second problem multiplied by a scalar and shifted by a constant. This second minimization problem defines the projection $P_C^G[x^k - wG^{-1}f(x^k)]$. Thus, the variational inequality subproblem of Step 1 is, in fact, equivalent to a projection.

Dafermos shows that the algorithm converges if the following assumptions are satisfied:

(1) f is continuously differentiable and strongly monotone;

(2) C is a polytope;

(3) the constant w satisfies $0 < w < 2c/k$, where c satisfies $(x-y)^T[f(x) - f(y)] > c\,||x-y||^2$ for every $x, y \in C$, and k is the maximum over C of the maximum eigenvalue of the positive definite symmetric matrix $[\nabla f(x)]^T G[\nabla f(x)]$.

The rate of convergence under these assumptions is linear with convergence ratio $[1 - (1/d)w(2c-kd)]^{1/2}$, where d is the maximum eigenvalue of G.

Dafermos suggests two possible choices for the matrix G: G a diagonal matrix or G "as close as possible" to the Jacobian matrix $\nabla f(x)$.

2.1.2 Bertsekas and Gafni: "Projection Methods for Variational Inequalities with Application to the Traffic Assignment Problem."

In this paper, the authors show that if the mapping $g: Y \subseteq R^m \to R^m$ and the set Y satisfy the three conditions stated at the beginning of this section that guarantee that the projection method can be used to find a solution y^* to VI(g,Y), then the projection method can be used to find the solution to the variational inequality problem VI(f,X), where $X \subseteq R^n$ is assumed polyhedral, $f = A^T gA : X \subseteq R^n \to R^n$, and A is a linear mapping from R^n into R^m.

This result is useful in the following situation. Suppose we are given the variational inequality problem VI(g(y),Y), where g is Lipschitz continuous and uniformly monotone. Instead of solving this problem over $y \in Y \subseteq R^m$, suppose that we would prefer to make a linear transformation of the problem and work in the space of $x \in X \subseteq R^n$, where

$$y = Ax \text{ and } Y = AX = \{y : y = Ax,\ x \in X\}.$$

That is, we would prefer to solve the equivalent variational inequality problem

$$(x-x^*)^T A^T g(Ax^*) \geq 0 \text{ for every } x \in X$$

or, equivalently,

$$(x-x^*)^T f(x^*) \geq 0 \text{ for every } x \in X,$$

where $f = A^T gA$.

The results in this paper allow us to use the projection method to solve the variational inequality problem in the space of x without making any explicit assumptions about f. Rather, g must meet the required conditions for convergence of the projection algorithm. (The mapping f need not inherit all of the convergence conditions imposed on g: In particular, the mapping $f = A^T gA$ need not be uniformly monotone unless $A^T A$ is nonsingular.)

The paper discusses this situation in the context of the traffic equilibrium problem. The projection algorithm devised by Dafermos [1980] for this problem operates in the space of link flows. The authors note that it is much less costly to carry out projection iterations in the space of path flows than in the space of link flows. The results in this paper allow the problem to be solved by the projection method in the space of path flows as long as the cost function defined on the links of the network meets the required convergence conditions.

The paper also presents a projection scheme that allows the matrix norm to be modified whenever the algorithm has made "reasonable progress" towards convergence.

2.1.3 Bakušinskiĭ and Poljak: "On the Solution of Variational Inequalities."

Most algorithms for solving the variational inequality problem assume that the mapping f is at least strictly, and usually uniformly, monotone over the ground set C. As we have seen, under either of these assumptions, the solution to the problem (if it exists) is unique. In this paper, the authors consider solving VI(f,C) when f is a monotone, Lipschitz continuous map. In this case, the solution to the problem is not guaranteed to be unique. The set M of solutions to VI(f,C) is, however, convex and closed.

The authors discuss projection algorithms for VI(f,C) when f is a point to set mapping from a Hilbert space into its dual space. The following discussion summarizes only the results for problems with single-valued maps. The projection algorithms in this paper treat explicitly a computation (or approximation) error in approximating $f(x^k)$, and allow the step size to vary at each iteration. Below we describe the three types of algorithms proposed in the paper: (1) a basic projection algorithm; (2) a regularization method; and (3) a method that alternately performs a regularization step and a projection step.

(1) The basic projection algorithm presented in the paper is

$$x^{k+1} = P_C[x^k - \alpha_k(f(x^k) + h_k)],$$

where P_C is the projection operator onto C, h_k is a vector of computation errors and α_k is the step length. The authors show that the algorithm converges under several different sets of assumptions. For example, the sequence x^k converges to the unique solution x^* if

(i) C is closed, convex and nonempty;

(ii) f is a single-valued, hemicontinuous, strongly monotone and Lipschitz continuous; and

(iii) the step size α_k and the error size $\|h_k\|$ satisfy several conditions.

(2) If f is monotone, the authors show that a regularization method can be used to construct a sequence of variational inequality problems, the solutions of which will converge to a point in the set M of solutions. For a given uniformly monotone regularizing operator $g:C \to R^n$, the k^{th} variational inequality subproblem is $VI(f^k,C)$, where $f^k(x) = f(x) + \varepsilon_k g(x) + h_k$, and $\varepsilon_k > 0$ for $k = 1,2,\ldots$ are regularization parameters that approach zero as $k \to \infty$.

The sequence $\{x^k\}$ converges to the unique point in M that solves VI(g,M) under the assumptions that

(i) C is closed, convex and nonempty, and

(ii) f is single-valued, hemicontinuous and monotone.

The authors note that Browder [1966] has developed a similar result.

(3) The authors propose an improvement of the regularization method which effectively combines the above projection and regularization methods and which obviates the need to solve explicitly the sequence of variational inequality subproblems generated by the regularization method. The sequence of iterates $\{x^k\}$ is given by

$$x^{k+1} = P_C[x^k - \alpha_k(f(x^k) + \varepsilon_k g(x^k) + h_k)],$$

where the regularization parameter decreases on each iteration. The sequence $\{x^k\}$ converges (as in (2)) to the unique solution to VI(g,M) under the assumptions of the regularization method and (for example) if:

(i) f and g are Lipschitz continuous with the same Lipschitz coefficient, and

(ii) α_k, ε_k, and $||h_k||$ satisfy several conditions.

2.2 Nonlinear Programming-Based Algorithms

We have noted earlier that if the Jacobian of f(x) is symmetric on the feasible region of the variational inequality problem, then f is integrable and the problem can be reformulated as an equivalent minimization problem and solved using an appropriate nonlinear programming algorithm. This fact motivates questions about whether nonlinear programming algorithms will converge when less restrictive conditions are imposed on f.

The following works define conditions under which certain nonlinear programming algorithms adapted to the variational inequality problem will converge.

2.2.1 Ahn and Hogan: "On Convergence of the PIES Algorithm for Computing Equilibria."

The authors analyze an iterative algorithm known as the PIES (Project Independence Evaluation System) algorithm to compute equilibria in economic equilibrium problems. When the market demand function is integrable, the competitive market equilibrium problem they consider is equivalent to an economic surplus maximization problem. The authors establish convergence criteria for models for which the supply mapping is monotone and the demand function is not necessarily integrable.

The main idea of the algorithm is to approximate the original market equilibrium problem (formulated as a fixed point problem) with a sequence of modified market equilibrium problems that have equivalent formulations as optimization problems. The modified problem at each iteration is constructed by "diagonalizing" the market demand function.

We can restate the PIES algorithm for the general variational inequality problem VI(f,C) as follows. Define an approximate mapping g(x,y) with component functions $g_i(x,y) = f_i(y_1,y_2,\ldots,y_{i-1},x_i,y_{i+1},\ldots,y_n)$ for $i = 1,2,\ldots,n$. Note that the i^{th} component function of g depends only on x_i. The algorithm can be stated as follows:

PIES Algorithm Adapted to the Variational Inequality Problem

Step 0: Select a feasible point $x^o \in C$. Set $k = 0$.

Step 1: Given $x^k \in C$, let x^{k+1} solve the variational inequality problem

$$(x-x^{k+1})^T g(x^{k+1},x^k) \geq 0 \text{ for every } x \text{ in } C.$$

If $x^{k+1} = x^k$, stop: $x^k = x^*$.

Otherwise, return to Step 1 with $k = k + 1$. ∎

The variational inequality subproblem in Step 1 has an equivalent formulation as an optimization problem because the Jacobian of $g(x,x^k)$ is diagonal, and, hence, symmetric. If C is a rectangular set ($C = C_1 x C_2 x \ldots x C_n$, where each $C_i \subseteq R^1$), then the problem is particularly easy to solve because it separates into n one-dimensional variational inequality problems of finding x_i^{k+1} satisfying $(x_i - x_i^{k+1})^T f_i(x_1^k,\ldots,x_i^{k+1},\ldots,x_n^k) \geq 0$ for every $x_i \in C_i$. Note that if f is integrable, the PIES algorithm is simply the nonlinear Jacobi method applied to the equivalent maximization problem.

The authors prove that the algorithm is globally convergent when the demand function is linear and the supply function is monotone, and that it is locally convergent when the demand function is nonlinear and the supply function is monotone. In terms of the variational inequality problem VI(f,C) local convergence results can be obtained under the following assumptions:

(i) C is convex;

(ii) a solution x^* exists;

(iii) f is differentiable on C and $\partial f_i/\partial x_i(x) \geq 0$ everywhere on C;

(iv) f is continuously differentiable in a neighborhood of x^*, and $\partial f_i/\partial x_i(x^*) > 0$ for every i; and

(v) $s = \|D^{-1/2}BD^{-1/2}\|_2 < 1$, where $D = \mathrm{diag}(\nabla f(x^*))$, $B = \nabla f(x^*) - D$.

Although this interpretation of the local convergence proof is not given in the paper, we can view the general idea of the proof in the following way. (See Chapter 4 for a formal description of this proof technique). The proof shows that if the initial iterate is sufficiently close to x^*, then for each x^k, $g(x,x^k)$ points away from x^*; i.e., $(x^*-x)^T g(x,x^k) < 0$ for every $x \in C$ satisfying $\|x-x^*\|_2 \geq r\,\|x^k-x^*\|_2$, where r is a given constant less than 1. Since any such x cannot solve the approximate variational inequality problem in Step 1; x^{k+1} must satisfy $\|x^{k+1}-x^*\|_2 < r\,\|x^k-x^*\|_2$, and the convergence of the sequence $\{x^k\}$ to x^* follows from the contraction mapping theorem.

2.2.2 Dafermos: "An Iterative Scheme for Variational Inequalities."

This work considers a general iterative scheme to solve the finite-dimensional variational inequality problem VI(f,C). The proposed scheme generalizes projection, linear approximation, and relaxation (Jacobi) methods. Each of these methods generates a sequence of iterates $\{x^k\}$ in C that, with suitable assumptions imposed on f(x), contract toward the solution x^* with respect to a given matrix norm. The general scheme given in this paper allows the matrix norm to vary at each iteration. Dafermos shows that the sequence of iterates $\{x^k\}$ generated by the algorithm satisfies $\|x^{k+1}-x^k\|_{k+1} \leq r\,\|x^k-x^{k-1}\|_k$ for some $r \in [0,1)$, where $\{\|\cdot\|_k\}$ is a sequence of norms in R^n induced by a sequence of symmetric positive definite matrices $\{G_k\}$.

Generalizing Ahn's nonlinear Jacobi method, the iterative scheme considered here defines a mapping $g(x,y)$ that approximates $f(x)$ about a point y so that (i) $g(x,x) = f(x)$ for every $x \in C$, and (ii) the variational inequality problem solved at the k^{th} iteration, namely,

$$\text{find } x^{k+1} \in C \text{ satisfying } (x-x^{k+1})^T g(x^{k+1},x^k) \geq 0 \text{ for every } x \in C,$$

is easy to solve.

This general algorithm converges if the following assumptions are satisfied:

(1) $C \subseteq R^n$ is compact and convex;

(2) $f:C \to R^n$ is continuously differentiable; and

(3) there exists a "smooth" mapping $g(x,y):CxC \to R^n$ satisfying

(i) $g(x,x) = f(x)$ for every x in C;

(ii) for every fixed $x \in C$ and $y \in C$, the $n \times n$ matrix $g_x(x,y)$ is symmetric and positive definite; and

(iii) $\|g_x^{-1/2}(x_1,y_1)g_y(x_2,y_2)g_x^{-1/2}(x_3,y_3)\|_2 < 1$ for every x_1,y_1,x_2,y_2,x_3,y_3 in C.

(Note: Dafermos shows that this norm condition on g implies that $f(x)$ is strictly monotone.)

The algorithm proceeds as follows.

Algorithm

Step 0: Select a feasible point $x^o \in C$. Set $k = 0$.

Step 1: Given $x^k \in C$, let x^{k+1} solve the variational inequality problem

$$(x-x^{k+1})^T g(x^{k+1},x^k) \geq 0 \text{ for every } x \text{ in } C.$$

Return to Step 1 with $k = k + 1$. ∎

Since $g_x(x,y)$ is a symmetric positive definite matrix for every $x \varepsilon C$ and $y \varepsilon C$, $g(x,y)$ is the gradient with respect to x of a strictly convex (in x) functional $F(x,y)$, i.e., $g(x,y) = F_x(x,y)$. The k^{th} variational inequality subproblem is, therefore, equivalent to the convex minimization problem

$$\min_{x \varepsilon C} F(x,x^k),$$

and can be solved using nonlinear programming techniques.

The algorithm reduces to the nonlinear Jacobi method (see Ahn and Hogan [1982], for example), if we define $g(x,y) = f_i(y_1,\ldots,y_{i-1},x_i,y_{i+1},\ldots,y_n)$.

The algorithm reduces to the nonlinear Gauss-Seidel method (see Pang and Chan [1982], for example) if we define $g(x,y)$ by $g_i(x,y) = f_i(x_1,x_2,\ldots,x_i,y_{i+1},\ldots,y_n)$. In this case, the requirement that the matrix $g_x(x,y)$ be symmetric is clearly too stringent, since this matrix is lower triangular. In fact, the assumption that $g_x(x,y)$ is symmetric is not necessary for the proof of the algorithm. This symmetry assumption does, however, ensure that the variational inequality problem subproblem of Step 1 can be solved as an equivalent convex minimization problem.

If $g(x,y) = f(y) + (1/w)G(x-y)$, then the algorithm reduces to the projection method

$$x^{k+1} = P_C^G(x^k - wG^{-1}x^k),$$

where $P_C^G(z)$ is the projection of the point z onto the set C with respect to the G norm and w is a positive "steplength." (See Section 2.1)

The algorithm reduces to the linear approximation method: x^{k+1} solves the variational inequality problem

$$(x-x^{k+1})^T f^k(x^{k+1}) \geq 0 \text{ where } f^k(x) = f(x^k) + A(x^k)(x-x^k)$$

if $g(x,y) = f(y) + A(y)(x-y)$, where $A(y)$ is an $n \times n$ matrix which depends continuously on y. (See, for example, Pang and Chan [1981].) This method is a Newton method when $A(y) = \nabla f(y)$.

2.2.3 Florian and Spiess: "The Convergence of Diagonalization Algorithms for Asymmetric Network Equilibrium Problems."

In this paper, the authors recast the nonlinear Jacobi method studied by Ahn [1979] in the traffic equilibrium setting. In this context, $f_i(x)$ represents the cost incurred by a single vehicle traversing arc i when x is the arc flow pattern on the network.

The "diagonalization" algorithm presented is an adaptation of the PIES algorithm to the variational inequality problem as presented in Section 2.2.1, and the proof of convergence is demonstrated under the assumptions stated in that section.

The authors note that when this diagonalization algorithm is used to solve the multimodal fixed demand network equilibrium problem, the variational inequality subproblems decompose by mode.

2.2.4 Pang and Chan: "Iterative Methods for Variational and Complementarity Problems."

This work studies the family of iterative linear approximation methods to solve the variational inequality problem VI(f,C). These methods generate a sequence of iterates by solving linear variational inequality subproblems. The k^{th} iteration of the algorithm approximates $f(x)$ in a neighborhood of x^k with a linear mapping. The solution of the variational inequality problem over C with this approximate mapping is the $(k+1)^{st}$ iterate, x^{k+1}. The algorithm can be stated as follows:

Linear Approximation Algorithm

Step 0: Find a feasible point $x^o \in C$. Set $k = 0$.

Step 1: Given $x^k \in C$, let $f^k(x)$ be the linear mapping

$$f^k(x) = f(x^k) + A(x^k)(x-x^k),$$

where $A(x^k)$ is an $n \times n$ matrix depending continuously on x^k.

Step 2: Let x^{k+1} solve the linear variational inequality problem:

$$(x-x^{k+1})^T f^k(x^{k+1}) \geq 0 \text{ for every } x \in C.$$

Return to Step 1 with $k = k + 1$. ∎

A number of well-known methods can be implemented as linear approximation methods:

(1) Newton Methods: $f(x)$ is assumed differentiable and $A(x^k) = \nabla f(x^k)$;

(2) Quasi-Newton Methods: $A(x^k)$ is an approximation to $\nabla f(x^k)$; and

(3) Projection Methods: $A(x^k) = G$ for every k, where G is a fixed positive symmetric $n \times n$ matrix. In this case, the solution x^{k+1} to the k^{th} linear subproblem is the projection onto the set C of the point $x^k - G^{-1}f(x^k)$ with respect to the G norm.

Let $L(x)$, $U(x)$, and $D(x)$ be respectively the strictly lower, strictly upper, and diagonal parts of the matrix $\nabla f(x)$.

(4) Linearized Jacobi Method: $A(x^k) = D(x^k)$.

(5) Linearized Gauss-Seidel Method: $A(x^k) = L(x^k) + D(x^k)$ or $A(x^k) = U(x^k) + D(x^k)$;

(6) SOR (Successive Overrelaxation) Methods: $A(x^k) = L(x^k) + (1/w)D(x^k)$ or $A(x^k) = U(x^k) + (1/w)D(x^k)$, where $w \in (0,2)$ is a relaxation parameter.

The authors prove local and global convergence under several different sets of conditions. The main result assumes the following assumptions for local convergence:

(1) C is closed and convex;

(2) f(x) is continuous;

(3) A(x) is continuous;

(4) there exists a positive definite matrix G satisfying

(i) $A(x^*) - G$ is positive semi-definite (and, hence, since G is positive definite, so is A);

(ii) $\|\hat{G}^{-1}[f(x) - f(y) - A(y)(x-y)]\|_{\hat{G}} \leq b \; \|x-y\|_{\hat{G}}$ for some $b < 1$ (i.e., the linear approximation is "good" in some neighborhood of x^* with respect to the G-norm) for all x,y in some neighborhood of x^*.

Global convergence of the linear approximation algorithm follows from conditions (1) and (2) above and the conditions

(3') A(x) is bounded on compact subsets of C; and

(4') $\|\hat{G}^{-1}[f(x) - f(y) - A(y)(x-y)]\|_{\hat{G}} < b\|x-y\|_{\hat{G}}$ for some $b < 1$ for all x, y in C (i.e., the linear approximation is "good" everywhere on C).

The paper also discusses the nonlinear Jacobi method. The assumptions and the proof of convergence are very similar to those given in Ahn's study of the PIES algorithm. In particular, both papers assume that f satisfies the norm condition $\|D^{-1/2}BD^{-1/2}\|_2 < 1$, where $D > 0$ and B are, respectively, the diagonal and off-diagonal parts of $\nabla f(x^*)$. Pang and Chan specify a sufficient condition for this norm condition to hold. The condition is particularly easy to verify: if $\nabla f(x^*)$ is row diagonally dominant and strictly column diagonally dominant (or column diagonally dominant and strictly row diagonally dominant) and has positive diagonal entries, then $||D^{-1/2}BD^{-1/2}||_2 < 1$.

Although the authors do not discuss methods to solve the linear variational inequality subproblem encountered in Step 2 of the algorithm, they note that there are many efficient algorithms to solve linear variational inequality problems. In particular, if C is polyhedral, the linear variational inequality problem can be reformulated as a linear complementarity problem and solved using any algorithm for the linear complementarity problem. They mention that Eaves [1978b,c] and Pang [1981] have recently developed algorithms specifically designed to solve linear variational inequality problems over polyhedral sets. Aashtiani and Magnanti [1982] report computational results for algorithms for this problem.

The authors state a number of other sets of conditions that imply convergence of the linear approximation algorithm. Some of these conditions are particularly well-suited for showing convergence of some of the methods (e.g., Newton's, Quasi-Newton's, linearized Jacobi) mentioned above. For example, local convergence of Newton's method follows under conditions (1) and (2) above and the condition that $\nabla f(x^*)$ is positive definite. Moreover, if $f(x^*)$ is Lipschitz continuous at x^*, then the sequence $\{x^k\}$ generated by Newton's method converges quadratically to the solution x^*.

2.2.5 Pang and Chan: "Gauss-Seidel Methods for Variational Inequality Problems over Product Sets."

This paper considers linear and nonlinear Gauss-Seidel type iterative methods to solve the variational inequality problem VI(f,C), where

$C \subseteq R^n$ is a product set; i.e., $C = C_1 x C_2 x \ldots x C_m$, where $C_i \subseteq R^{n_i}$ and $n_i \geq 1$ for $i = 1,\ldots,m$, and $\sum_{i=1}^{m} n_i = n$.

The nonlinear Gauss-Seidel method generates a sequence of iterates $\{x^k\}$ as follows:

Given $x^k = (x_1^k,\ldots,x_m^k)$ and x_i^{k+1} for $i < j \leq n$ (where $x_i \varepsilon C_i \subseteq R^{n_i}$), let x_j^{k+1} solve the following nonlinear variational inequality problem over C_j:

$$(x_j - x_j^{k+1})^T f_j(x^{k,j}) \geq 0 \text{ for every } x_j \varepsilon C_j,$$

where $x^{k,j} = (x_1^{k+1},\ldots,x_j^{k+1},x_{j+1}^k,\ldots,x_m^k)$.

The algorithm is locally convergent if

(1) C_i is nonempty, closed, and convex for $i = 1,\ldots,m$;

(2) f is continuously differentiable;

(3) a solution x^* to the variational inequality problem VI(f,C) exists; and

(4) there exist positive definite matrices G_i for $i = 1,\ldots,m$ such that $\nabla_i f_i(x^*) - G_i$ is positive semi-definite and $\|G^{-\frac{1}{2}}BG^{-\frac{1}{2}}\| < 1$, where

$$G = \begin{bmatrix} G_1 & & 0 \\ & G_2 & \\ & & \ddots \\ 0 & & G_m \end{bmatrix} \text{ and } B = \nabla f(x) - \begin{bmatrix} \nabla_1 f_1(x^*) & & 0 \\ & \nabla_2 f_2(x^*) & \\ & & \ddots \\ 0 & & \nabla_m f_m(x^*) \end{bmatrix}$$

and the matrix norm $\|A\|$ is the "block-∞" norm induced by the "block-∞" vector norm defined by

$$\|x\| = \text{Max} \|x_i\|_2 \ .$$

That is, $||A|| = \max_{||x||=1} ||Ax|| = \max_{1 \le i \le m} \max_{||x||=1} ||(Ax)_i||_2$.

If we assume that each C_i is contained in R^1, then condition (4) reduces to $D > 0$ and $\|D^{-1/2}BD^{-1/2}\|_\infty < 1$, where $D = \text{diag}\, \nabla f(x^*)$, $B = \nabla f(x^*) - D$, and $\|\cdot\|_\infty$ is the usual matrix ∞-norm.

The proof of the nonlinear Gauss-Seidel method uses arguments similar to those in Ahn's convergence proof of the nonlinear Jacobi method.

The linear Gauss-Seidel method generates a sequence of iterates x^k as follows:

Given x^k and x_i^{k+1} for $i \le j < n$, let x_{j+1}^{k+1} solve the linear variational inequality problem over C_{j+1}:

$$(x_j - x_{j+1}^{k+1})^T[(f_j(x^{k,j}) + A_j(x^{k,j})(x_{j+1}^{k+1} - x_{j+1}^k)] \ge 0$$

where $A_i(x)$ is an $n_j x n_j$ matrix depending continuously on x and $x^{k,j}$ is defined above.

Results similar to those in Pang and Chan [1981] are given for various choices of the mapping A_j.

2.2.6 Rockafellar: "Lagrange Multipliers and Variational Inequalities."

In this paper, the author shows that by "discretizing" the ground set C (i.e., approximating C by a finite set of convex inequalities) and introducing Lagrange multipliers, VI(f,C) can be formulated as a generalized set of Kuhn-Tucker optimality conditions, and, hence, as a nonlinear complementarity problem. Both of these formulations suggest computational procedures to solve the problem. This paper, for example, focuses on penalty-duality methods to find a solution satisfying the generalized Kuhn-

Tucker conditions for the problem. (Auslender [1976] and Bakusinskii [1979] also discuss penalty methods for this formulation.)

In order to establish the optimality conditions for the problem, we approximate the ground set C with the finite set of constraints $C_0 := \{x \in R^n : g_1(x) \leq 0, \ldots, g_m(x) \leq 0\}$, where each g_i is a convex, differentiable function. Let $g(x) = (g_1(x), \ldots, g_m(x))$ and assume that some point $x \in R^n$ satisfies $g(x) < 0$. Then x^* is a solution to $VI(f, C_0)$ if and only if $-f(x^*)$ is in the normal cone to C_0 at x^*, which is true if and only if $-f(x^*)$ is a positive linear combination of the gradients $\nabla g_i(x^*)$ of the active constraints at x^*; i.e., if and only if $x^* \in R^n$ and there exists a $u^* \in R^m$ satisfying

$$f(x^*) + \nabla g(x^*)u^* = 0,$$

$$g(x^*) \leq 0,$$

$$u^* \geq 0,$$

and

$$(u^*)^T g(x^*) = 0.$$

The author describes an extension of the Hestenes-Powell penalty-duality method to solve this system of inequalities. At each iteration, the method must find an unconstrained zero of an augmented Lagrangian function. A simple updating rule uses the solution to this subproblem to produce the next iterate. The problem with using this method to solve variational inequality problems (as opposed to optimization problems) is that the convergence proof assumes that the zero to the unconstrained augmented Lagrangian found at each iteration is an exact solution to the subproblem.

The author proposes using the proximal method of multipliers, a modified version of the Hestenes-Powell algorithm, to solve VI(f,C_0). In this case, the algorithm must find an approximate unconstrained zero of an augmented Lagrangian at each iteration. (The subproblem reduces to finding the proximation of the Hestenes-Powell augmented Lagrangian if f is a gradient mapping.) The rate of convergence of the method is linear for "almost all" variational inequality problems. Luque [1984] analyzes the asymptotic rate of convergence of proximal point algorithms.

Rockafellar also shows that the generalized Kuhn-Tucker optimality conditions for VI(f,C_0) can be formulated as a nonlinear complementarity problem. He shows that if f is monotone and continuous and each g_i is convex, then the mapping of the nonlinear complementarity problem is monotone and continuous relative to the nonnegative orthant. However, the map of the complementarity problem cannot be uniformly monotone (even if f is) unless the problem is unconstrained, and cannot be the gradient of any function, even if f is. As Rockafellar points out, this result indicates a natural way in which nonlinear complementarity problems (and, hence, variational inequality problems) with neither uniformly monotone nor gradient-type maps arise in optimization theory.

2.2.7 Luthi: "On the Solution of Variational Inequalities by the Ellipsoid Method."

This paper extends the ellipsoid method of linear programming (Khachiyan [1979]) to solve VI(f,C). It assumes that the ground set C is a nonempty, closed, convex subset of R^n and that f is continuous and monotone. Under these assumptions, the solution set to VI(f,C) is the set

$$M = \{x^* \varepsilon C: (x - x^*)^T f(x) \geq 0 \text{ for every } x \varepsilon C\}$$

$$= \bigcap_{x \varepsilon C} \{x^* \varepsilon C: (x - x^*)^T f(x) \geq 0\},$$

which is closed and convex. (Note that this formulation replaces $f(x^*)$ in the variational inequality problem with $f(x)$. This replacement is valid for problems with continuous, monotone maps: see Section 2.4.)

Relying on this reformulation, Luthi describes the ellipsoid algorithm for VI(f,C) as follows:

> We start with the convex body C included in a ball $S(a_0,R) = E_0$ and the monotone continuous mapping f. In the k-th step there will be a smaller set C_k of C which includes the solution set to V(f,C) and is contained in the ellipsoid E_k with center x_k. If its center $x_k \notin C$, then we generate a hyperplane through x_k which avoids C. This hyperplane cuts E_k into two halves. We pick the [half including] C_k and include it in a new ellipsoid E_{k+1}, "smaller" than E_K. In the case $x_k \varepsilon C$, we cut with the hyperplane $f(x_k)^T x_k = f(x_k)^T x$ similarly and take the part with $f(x_k)^T x_k \geq f(x_k)^T x$ to be included in the new ellipsoid. ...
> The volumes of the ellipsoids E_k will tend to 0 exponentially and this, with the assumption of strong monotonicity, will guarantee that a subsequence of these centers x_k which are in C will tend to a solution exponentially fast.

When the ellipsoid method is used to solve linear programs, a non-degeneracy assumption must be imposed on the problem to ensure that the ellipsoids do not converge to a hyperplane by becoming "flatter and flatter." When the algorithm is extended to solve VI(f,C), imposing a uniform monotonicity condition on f avoids this problem.

2.3 Simplicial Decomposition Algorithms

A general simplicial decomposition algorithm (see Von Hohenbalken [1977], for example) is an iterative method that alternatively solves a master problem and a subproblem. The subproblem generates (in some cases affinely independent) points contained in the feasible region of the original problem; the master problem solves a restricted version of the original problem over the convex hull of these "generator" points and may reduce the set of feasible points generated by the previous subproblems by dropping points that are not needed to represent the current solution of the restricted master problem. In the general framework developed by Geoffrion [1971], the subproblem represents the "inner linearization" of the feasible region (the approximation of the feasible region by the convex hull of points contained in the feasible region); the master problem represents a "restriction" of the original problem (the original problem is solved over a restricted subset of the feasible region). The master problem is usually expressed in terms of the barycentric coordinates of the set of generator points that define the restricted feasible region.

The following works describe algorithms that belong to the general class of simplicial decomposition algorithms.

2.3.1 Smith: "The Existence and Calculation of Traffic Equilibria" and "An Algorithm for Solving Asymmetric Equilibrium Problems with a Continuous Cost-Flow Function."

In these companion papers, Smith develops a simplicial decomposition algorithm to solve the traffic equilibrium problem and proposes an algrithm to solve the restricted master problem. Smith's algorithm consists of an outer algorithm that generates an extreme point of the feasible region at each iteration, and an inner algorithm that solves the variational inequality problem over the convex hull of the previously generated extreme points.

General Algorithm

Outer Algorithm (Generates extreme points)

Step 0: Select $x^0 \varepsilon C$.

Step 1: Given, $x^k \varepsilon C$, let v^k be a solution to the problem

$$\min_{x \varepsilon C} (x - x^k)^T f(x^k)$$

If $(x^k)^T f(x^k) = (v^k)^T f(x^k)$, stop with $x^k = x^*$.

Otherwise, go to Step 2.

Inner Algorithm (Solves a variational inequality problem over a restricted feasible region).

Step 2: Let x^{k+1} solve the variational inequality problem over the convex hull of $v^1,\ldots,v^k$; i.e., x^{k+1} satisfies $(v^i - x^{k+1})^T f(x^{k+1}) \geq 0$ for $i = 1,\ldots,k$.

Go to Step 1 with k=k+1. ∎

If C is assumed polyhedral, the minimization problem of Step 1 is a linear program. Smith's major contribution in these papers is a method for solving the restricted subproblem of Step 2. The inner algorithm is capable of solving a variational inequality problem over a feasible region whose extreme points are all known. It uses an objective function $V(x^k)$ that measures the "departure from equilibrium" (i.e., the distance from the solution x^*) of a point x^k in the feasible region, and defines a direction of descent Δ_k from x^k for the function $V(x)$. Because the algorithm requires that all extreme points of the feasible region are known, which is unlikely to be met in practice, it would be useful only in conjunction with an extreme point generating scheme such as that presented here.

Smith formulates the problem over a feasible set $C = S \cap D$ defined by an open, convex, bounded set $S \subseteq R^n$ and a closed and convex set $D \subseteq R^n$. If $D \subseteq S$, no further conditions on the feasible region are necessary. If $D \not\subseteq S$, a coercivity-type boundary condition is required to ensure that a solution exists. If we assume that the variational inequality problem is defined over a convex and compact ground set C, then the results require no boundary condition. The conditions for convergence are, therefore:

(i) C is convex and compact; and

(ii) f is continuously differentiable and monotone on C. (Note that <u>strict</u> monotonicity is not required).

Given a finite set of extreme points $\{v^i\}$, the function

$$V(x) = \sum_{i=1}^{k} ([(x - v^i)^T f(x)]^+)^2,$$

where $[c]^+ = \mathrm{Max}\{0,c\}$, measures the departure from equilibrium of any point x in the convex hull of the v^i. Note that $V(x) \geq 0$ for every $x \in C$, and $V(x) = 0$ if and only if x is an equilibrium point in the convex hull of the v^i. Let $\Delta: R^n \to R^n$ be the direction given by

$$\Delta(x) = \frac{\sum_{i=1}^{k} ([(x-v^i)^T f(x)]^+ (v^i - x)}{\sum_{i=1}^{k} [(x - v^i)^T f(x)]^+} = \sum_{i=1}^{k} \alpha_i (v^i - x) \text{ where } \alpha_i \geq 0, \text{ and } \sum_{i=1}^{k} \alpha_i = 1.$$

If f is monotone, then the directional derivative of V(x) in the direction $\Delta(x)$ is negative, because

$$\Delta(x)\nabla V(x) \leq -2V(x),$$

and $V(x) < 0$ if x is not an equilibrium point. Therefore, $\Delta(x)$ is a descent direction for $V(x)$ from any nonoptimal point x.

Algorithm (Used to solve the inner problem.)

Step 0: Select $x^0 \varepsilon C$. Let $\delta_0 = 0$. Set $i = 0$.

Step 1: Given (x^i, δ_i), where $x^i \varepsilon C$ and $\delta_i \varepsilon (0,1]$;

if $V(x^i + \delta_i \Delta(x^i)) \leq (1 - \delta_i)V(x^i)$,

$$\text{then} \begin{cases} x^{i+1} = x^i + \delta_i \Delta(x^i) \text{ and} \\ \delta_{i+1} = \delta_i; \end{cases}$$

if $V(x^i + \delta_i \Delta(x^i)) > (1 - \delta_i)V(x^i)$,

$$\text{then} \begin{cases} x^{i+1} = x^i \text{ and} \\ \delta_{i+1} = \frac{1}{2}\delta_i. \end{cases}$$

Return to Step 1 with $i = i + 1$. ∎

These papers also prove convergence for this algorithm when the restricted subproblem is approximately solved. This property is of practical importance because it is unrealistic to assume that the restricted variational inequality problem could be solved exactly at each iteration.

2.3.2 Lawphongpanich and Hearn: "Simplicial Decomposition of the Asymmetric Traffic Assignment Problem"

The general framework of the algorithm presented in this paper is similar to that of Smith's [1983a,b] algorithm (described in the last subsection). In general terms, at each iteration, the algorithm

(1) approximately solves a "master" variational inequality problem over the convex hull of the current set of "generator points", and

(2) solves a linear programming subproblem. If the solution from (1) is not optimal, the algorithm adds this solution to the current set of generator points. Furthermore, if the "gap function" (as defined below) evaluated at the current linear program solution is sufficiently smaller than the gap function evaluated at all previous linear program solutions, the procedure drops all generator points with zero weight in the expression of the current linear program solution.

Let f be a monotone, continuous map. The gap function $G(x)$, used to measure the departure from equilibrium of the point $x \varepsilon C$, is defined as

$$G(x) = (x - y_x)^T f(x),$$

where y_x solves the linear program:

$$\min_{y \varepsilon C} y^T f(x).$$

From the definition of y_x, $G(x) \geq 0$ for every $x \varepsilon C$; moreover, $G(x) = 0$ if and only if x solves the variational inequality problem.

The authors suggest solving the restricted master variational inequality problem using the projection method of Bertsekas and Gafni [1980] described earlier. This method allows the linear transformation of the variables from $x \in C$ to $z \in Z$, where $x = Az$ and $Z = \{z: \sum_{i=1}^{m} z_i = 1$ and $z_i \geq 0$ for $i = 1,\ldots,m\}$. Here, the z_i give a coordinate representation of the vector x in terms of the extreme points. (A is an nxm matrix whose columns are the m extreme points of C.) If f(x) is Lipschitz continuous and strongly monotone in the space of x, then Bertsekas and Gafni's projection algorithm will solve the variational inequality problem in the coordinate space of z.

Computational results in the paper compare this algorithm with that of Nguyen and Dupuis [1981] and Bertsekas and Gafni [1980] for four test problems. The results are promising: the algorithm tends to generate and retain very few extreme flow patterns.

2.3.3 Pang and Yu: "Linearized Simplicial Decomposition Methods for Computing Traffic Equilibria of Networks"

This paper describes a simplicial decomposition method to solve the finite-dimensional variational inequality formulation of the traffic equilibrium problem. The general scheme is as follows:

Each iteration of the algorithm

(1) solves a linearized version of the variational inequality problem over the convex hull of a set of "generator" extreme points of the feasible region; and

(2) solves a linear program to test for optimality. If the solution from (1) is not optimal, the procedure adds the solution of the linear program to the current set of generator extreme points. (Certain points may also be dropped from the current set of generator points.)

The algorithm is a modified version of the Dantzig-van de Panne-Whinston algorithm (van de Panne and Whinston [1969]) for quadratic programming problems. (Pang [1981] shows that using this algorithm to solve a quadratic program produces the same sequence of primal feasible flow vectors as the sequence produced by Von Hohenbalken's simplicial decomposition algorithm [1977] specialized to the same quadratic programming problem.) Placing this algorithm in the simplicial decomposition framework eliminates the need to compute shortest paths, which may be time consuming.

The authors prove local and global convergence for the algorithm and provide extensive computational results. These results indicate that the method is quite promising for solving large-scale equilibrium problems.

2.4 Algorithms Based on Reformulating the Variational Inequality Problem as a Max-Min Problem

If f is monotone and hemicontinuous, then VI(f,C) has a number of equivalent formulations. In particular, VI(f,C) is equivalent to two "max-min" problems. Thus, even if VI(f,C) is not equivalent to a convex minimization problem in the "usual" way discussed in Section 1.3, it still is possible to reformulate the problem as a nonlinear programming problem.

Theorem 2.1 (Auslender [1976]).

Let C be closed, convex and nonempty. If f is monotone and hemicontinuous on C, then the following problems are equivalent:

Find $x^* \varepsilon C$ satisfying $(x - x^*)^T f(x^*) \geq 0$ for every $x \varepsilon C$. VI(f,C)

Find $x^* \varepsilon C$ satisfying $(x - x^*)^T f(x) \geq 0$ for every $x \varepsilon C$. VI'(f,C)

Find the solution x^* to $\max\limits_{x \varepsilon C} \min\limits_{y \varepsilon C} (y-x)^T f(x)$. MM(f,C)

Find the solution x^* to $\max\limits_{x \varepsilon C} \min\limits_{y \varepsilon C} (y - x)^T f(y)$. MM'(f,C)

(Hearn, Lawphongpanich and Nguyen [1983] discuss a "duality relationship" between the two max-min formulations.)

<u>Corollary</u>

Let C and f satisfy the conditions specified in Theorem 2.1. Then the set M of solutions to VI(f,C) is closed and convex (and possibly empty).

Let $g_y(x) = (y - x)^T f(x)$, and $G(x) = \min_{y \in C} g_y(x)$. Then MM(f,C) is equivalent to the nonlinear maximization problem

$$\max_{x \in C} \; G(x)$$

The function G(x) is, in general, not concave. For a given x, determining G(x) requires minimizing a linear objective function over C: this is a linear program if C is polyhedral.

Let $h_y(x) = (y - x)^T f(y)$, and $H(x) = \min_{y \in C} h_y(x)$. Then MM'(f,C) is equivalent to the nonlinear maximization problem

$$\max_{x \in C} H(x).$$

Since H(x) is the pointwise minimum of functions that are linear in x, it is concave. Thus, MM'(f,C) is a convex programming problem. If f is affine, determining H(x) for a given x requires minimizing a convex quadratic function over C. In general, the function $h_y(x)$ is not convex.

The papers in this section consider using nonlinear programming algorithms to solve the variational inequality problem stated in a max-min formulation. The first two papers discuss algorithms that approximate

$H(x)$ by the minimum of a finite number of functions at each iteration. The third paper discusses algorithms to solve the nondifferentiable, nonconvex minimization problem of minimizing $-G(x)$ over C. The last paper describes two descent methods: one for solving an unconstrained variational inequality problem, and one for solving a constrained variational inequality problem by maximizing $G(x)$ over C.

2.4.1 Auslender: Generalization of the Zuhovickii-Polyak-Primack (ZPP) Method

Zuhovickii, Polyak and Primack [1969] developed a method for solving n-person games. Auslender generalizes this method to solve VI(f,C) as follows:

Algorithm

Step 0: Select a feasible point $x^0 \varepsilon C$. Set $k = 0$.

Step 1: Given $x^1, \ldots x^k$, let x^{k+1} solve

$$\max_{x \varepsilon C} H_k(x),$$

where

$$H_k(x) = \operatorname{Min}\{(x^i - x)^T f(x^i) : i = 0,1,\ldots k\}.$$

If $x^{k+1} = x^i$ for some $i \leq k$, stop: x^{k+1} solves the variational inequality problem.

Otherwise, return to Step 1 with $k = k+1$. ∎

Auslender shows that if

(1) f is continuous and uniformly monotone on C, and

(2) C is convex and compact,

then this algorithm either terminates after a finite number of steps with $x^N = x^*$, or it generates an infinite sequence $\{x^k\}$ containing a subsequence that converges to x^*.

The subproblem

$$\underset{x \varepsilon C}{\text{Max}}\ H_k(x)$$

is a convex programming problem with a piecewise linear objective function. The problem can be reformulated with a linear objective function as follows:

$$\begin{aligned} &\text{Max } z \\ &\text{subject to: } (x^i - x)^T f(x^i) \geq z \qquad i = 0,1,\ldots,k \\ &\qquad\qquad x \varepsilon C. \end{aligned}$$

If C is polyhedral, the subproblem is a linear program.

To interpret the method geometrically, consider the max-min formulation of VI(f,C) introduced earlier:

$$\underset{x \varepsilon C}{\text{Max}}\ \underset{y \varepsilon C}{\text{Min}}\ (y - x)^T f(y). \qquad \text{MM}'(f,C)$$

As noted in Theorem 2.1, if f is a monotone hemicontinuous mapping, then x^* solves MM'(f,C) if and only if x^* solves VI(f,C). Letting $H(x) = \underset{y \varepsilon C}{\text{Min}}\ (y - x)^T f(y)$, we can consider $H_k(x)$ as a piecewise linear approximation to $H(x)$. As illustrated in Figure 2.2, this approximation is not obtained by making a tangent plane approximation to $H(x)$ at each x^k. Instead, each plane $r = (x^k - x)^T f(x^k)$ passes through the point $x = x^k$ and $r = 0$.

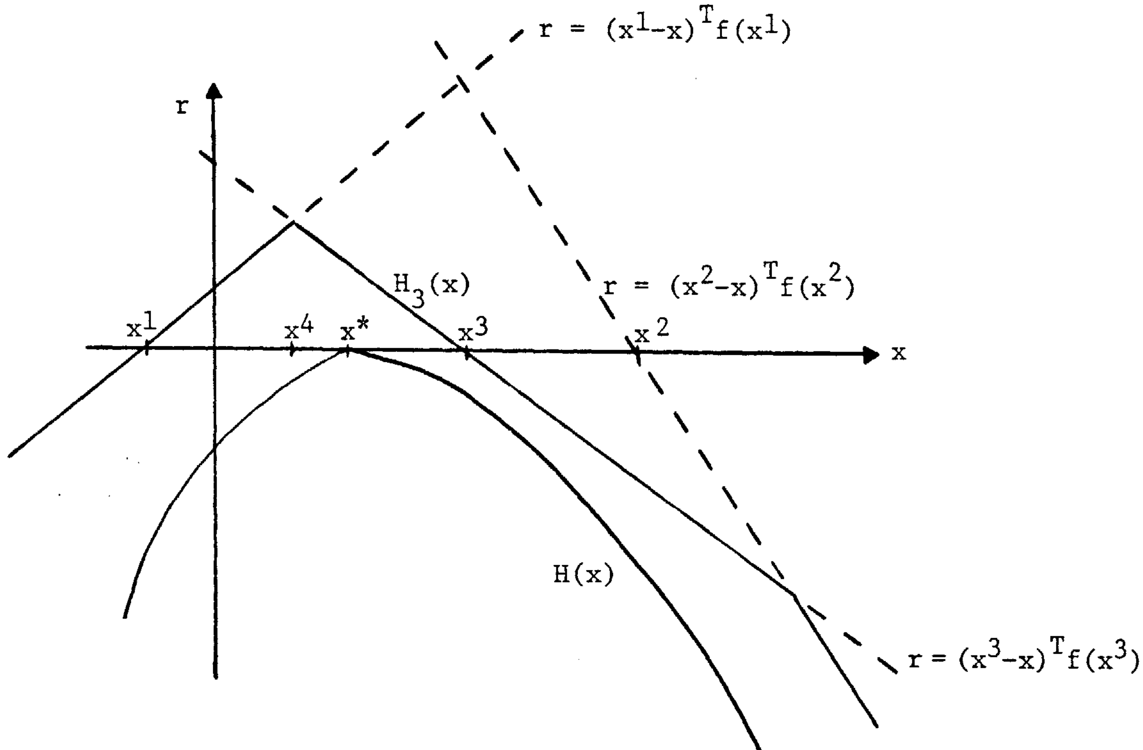


Figure 2.2

$H_k(x)$ is a Piecewise Linear Approximation to $H(x)$

2.4.2 Nguyen and Dupuis: "Un Méthode Efficace de Calcul d'un Trafic D'Équilibre dans le Cas des Coûts Non-symétriques."

The authors present a cutting plane algorithm to solve the variational inequality problem VI(f,C), and show that the algorithm fits into the framework of Zangwill's [1969] general cutting plane method. Our discussion will reinterpret this algorithm as a modification of Auslender's generalization of the ZPP method. On the k^{th} iteration, Auslender's algorithm moves to the solution of the subproblem $\{\text{Max } H_k(x): x \in C\}$. In contrast, this modified procedure moves to the point on the line segment $[x^k, y^k]$, where x^k is the previous iterate and y^k is the solution of the subproblem, that solves a one-dimensional variational inequality problem on $[x^k, y^k]$. This modified algorithm can be stated as follows.

Algorithm

Step 0: Select a feasible point $x^0 \varepsilon C$. Set $k = 0$.

Step 1: Given $x^0, x^1, \ldots, x^k$, let y^k solve the problem:

$$\underset{x \varepsilon C}{\text{Max}}\ H_k(x),$$

where

$$H_k(x) = \text{Min}\{(x^i - x)^T f(x^i) : i = 0,1,\ldots,k\}.$$

If $H_k(y^k) \leq 0$, stop: $y^k = x^*$.

Otherwise, go to Step 2.

Step 2: Let w_k solve the one-dimensional variational inequality problem on the line segment $[x^k, y^k]$: $w_k \varepsilon [0,1]$ satisfies $[[(1 - w)x^k + wy^k] - [(1 - w_k)x^k + w_k y^k]]^T f[(1 - w_k)x^k + w_k y^k] \geq 0$ for every $w \varepsilon [0,1]$.
That is, $(w_k - w)(x^k - y^k)^T f[(1 - w_k)x^k + w_k y^k] \geq 0$ for every $w \varepsilon [0,1]$. Go to Step 1 with $x^{k+1} = (1 - w_k)x^k + w_k y^k$ and $k = k + 1$. ∎

Nguyen and Dupuis show that if

(1) f is continuous and strictly monotone on C, and

(2) C is a nonempty polytope,

then the algorithm either terminates in a finite number of steps with the solution x^*, or it generates an infinite sequence $\{x^k\}$ that contains a subsequence converging to x^*.

Since C is assumed to be polyhedral, the subproblem of Step 1 is a linear program. The authors use a combination of Dantzig-Wolfe decomposition and the dual simplex method to reoptimize the linear program from the k^{th} to the $k+1^{st}$ iteration. To solve the one-dimensional variational inequality problem in Step 2, they suggest using Newton's Method to calculate the root of the equation

$$(x^k - y^k)^T f[(1 - w_k)x^k + w_k y^k] = 0.$$

They note that their procedure is the only convergent algorithm under the hypothesis of continuity and strict monotonicity that does not require the resolution of a convex or quadratic subproblem.

This method appears to be quite promising. The linear program subproblem is easy to solve, yet contains a great deal of information. On the k^{th} iteration, the subproblem

$$\max_{x \in C} \; \min_{i=1,\dots k} \; (x^i - x)^T f(x^i)$$

considers all of the points previously generated, and the value of the function f at each of these points. (In contrast, the Frank-Wolfe method looks myopically at $f(x^k)$.) Although the maximization in this subproblem is over all $x \in C$, the candidates for the solution are restricted to the set

$$S_k = \{x \in C: (x^i - x)^T f(x^i) \geq 0 \quad \text{for } i = 0,1,\ldots,k\}.$$

Thus, the size of the feasible region decreases at each iteration.

2.4.3 Marcotte: "A New Algorithm for Solving Variational Inequalities over Polyhedra, with Application to the Traffic Assignment Problem"

The author reformulates VI(f,C) as the equivalent minimization problem

$$\min_{x \in C} g(x), \text{ where } g(x) = \max_{y \in C} (x - y)^T f(x).$$

The "gap" function g(x) is, in general, nondifferentiable and nonconvex. Marcotte develops two versions of an algorithm based on nondifferentiable optimization techniques to minimize g(x).

The algorithms rely on the following result. Let R(x) be the set of solutions to the linear program $\max_{y \in C}(x - y)^T f(x)$. Then, for any $x \neq x^*$ in C, there exists a convex combination of extreme points of R(x) such that the direction from x to that point is a descent direction for g(x). If C is polyhedral and all of the extreme points of R(x) are known, a linear programming subproblem can be used to find a linear combination of the extreme points which gives a "steepest" descent direction for g(x). If the extreme points are not all known, then they can be generated as required by the procedure.

2.4.4 Auslender: Descent Algorithm-Based Methods

Auslender has developed algorithms to solve VI(f,C) that are based on descent methods to solve nonlinear programs. Consider the nonlinear program

$$\min_{x \in C} F(x) \qquad \text{(NLP)}$$

where

(1) C is a nonempty closed convex subset of R^n;

(2) $F: C \subseteq R^n \to R^n$ is finite, continuous and inf-compact (i.e. the level sets $S_\alpha = \{x \in C | f(x) \leq \alpha\}$ are compact for every $\alpha \in R$) on C; and

(3) the directional derivative $F'(x;y-x)$ exists for all $x,y \in C$.

A necessary condition for $x \in C$ to solve this minimization problem is that x is a stationary point of the problem, i.e. that

$$F'(x;\ y - x) \geq 0 \text{ for every } y \in C.$$

(This condition is also sufficient if F is convex on C.)

The general framework for the descent methods to find stationary points of the problem (NLP) is as follows:

Descent Algorithm

Step 0: Select $x^0 \in C$. Set $k = 0$.

Step 1: Given $x^k \in C$, let $d_k \in D(x^k)$. (The point to set mapping D(x) specifies a set of feasible directions from x.)

Step 2: Given x^k and d_k, let $x^{k+1} = x^k + w_k d_k$, where $w_k \varepsilon T_k(x^k, d_k)$. The point to set mapping $T_k(x,d)$ specifies a set of steplengths from x in the direction d.)

Go to Step 1 with $k = k+1$. ∎

Two common choices for the steplength mapping T_k are

(1) $T_k = T$, where $T(x, d) = \{w^*: F(x + w^*d) = \min F(x + wd)$ subject to $w \geq 0$ and $x + wd \varepsilon C\}$.

(2) $T_k(x,d) = \lambda_k$, where $\lim \lambda_k \to 0$, $\sum_{k=1}^{\infty} \lambda_k = +\infty$, and $\lambda_k > 0$ for $k = 0,1,2,\ldots$

For the first choice of steplengths, the sequence $F(x^k)$ decreases at each iteration: this need not be true for the second choice.

Auslender states conditions under which this general algorithm produces a sequence of iterates $\{x^k\}$ containing a subsequence converging to a stationary point. He presents the following two algorithms, which are based on this general descent algorithm, to solve the variational inequality problem.

Unconstrained Problem

The first algorithm is appropriate for the unconstrained problem $VI(f,R^n)$, which is equivalent to the problem

$$\text{Find } x^* \varepsilon R^n \text{ satisfying } f(x^*) = 0.$$

Assume that f is continuously differentiable and uniformly monotone on R^n. If we define

$$G(x) = (1/2) \| f(x) \|^2 ,$$

then the uniform monotonicity of f ensures that G is inf-compact. The set of stationary points to the problem

$$\inf_{x \varepsilon R^n} G(x) \tag{P}$$

is the set of solutions to the unconstrained variational inequality problem. To establish this property, first note that G(x) is differentiable, with $G'(x) = \nabla f(x)^T f(x)$. x is a stationary point of the problem (P) if and only if $G'(x) = \nabla f(x)^T f(x) = 0$, which is true if and only if $f(x) = 0$ (since $\nabla f(x)$ is positive definite and, therefore, nonsingular). Thus, we may determine a solution to $VI(f,R^n)$ by finding a stationary point of (P).

Consider the descent algorithm stated above with $D(x) = -f(x)$ (the "steepest" descent direction) and T_k the minimizing steplength:

<u>Algorithm</u>

Step 0: Select $x^0 \varepsilon R^n$. Set $k = 0$.

Step 1: Given $x^k \varepsilon R^n$, let $x^{k+1} = x^k - w_k f(x^k)$ where w_k solves the one-dimensional minimization problem

$$\operatorname*{Min}_{w \geq 0} G(x^k - wf(x^k))$$

If $f(x^{k+1}) = 0$, stop: $x^{k+1} = x^*$.

Otherwise, return to Step 1 with $k = k + 1$. ∎

Auslender shows that if f is continuously differentiable and uniform monotone, then this descent algorithm determines a stationary point to the problem (P) and, hence, a solution to $VI(f,R^n)$.

Constrained Problem

The second algorithm is appropriate for the constrained variational inequality problems VI(f,C) under the following assumptions:

(1) f is continuously differentiable and uniformly monotone;

(2) C is compact and strongly convex (i.e., for every x, y in C with $x \neq y$ and every $\lambda \varepsilon (0,1)$, there exists a scalar $r > 0$ such that $z \varepsilon C$ whenever $\|\lambda x + (1-\lambda)y - z\| < r$); and

(3) no point in C satisfies $f(x) = 0$.

Consider the max-min formulation of VI(f,C):

$$\text{Find } x^* \varepsilon C \text{ satisfying } G(x^*) = \max_{x \varepsilon C} G(x), \qquad (MM(f,C))$$

where $G(x) = \min_{y \varepsilon C} (y-x)^T f(x)$. Because $f(x) \neq 0$ on C, the strong convexity of C ensures that for a given x, the minimization problem

$$\min_{y \varepsilon C} (y - x)^T f(x) \qquad (P_{min}(x^*))$$

has a unique solution y_x, and, hence, that G is differentiable, with $G'(x) = (y_x - x)^T \nabla f(x) - f^T(x)$. The set of stationary points to the maximization problem

$$\max_{x \varepsilon C} G(x) \qquad (P_{max})$$

consists of those $x \in C$ satisfying the following inequality for every $y \in C$:

$$G'(x;y-x) = G'(x)(y-x)$$
$$= (y_x-x)^T \nabla f(x)(y-x)-(y-x)^T f(x) \leq 0.$$

The only stationary point to P_{max} is the unique solution x^* to VI(f,C). To establish this assertion, first note that if x^* solves VI(f,C), then x^* solves the minimization problem $P_{min}(x^*)$. Thus, $G'(x^*,y-x^*) = (x^*- x^*)^T \nabla f(x^*)(y-x^*)-(y-x^*)^T f(x^*) = -(y-x^*)^T f(x^*) > 0$ for every $y \in C$. Conversely, if x is a stationary point to P_{max}, then

$$0 \leq (y_x-x)^T \nabla f(x)(y_x-x) \leq (y_x-x)^T f(x) = \min_{y \in C} \quad (y-x)^T f(x) \leq (y-x)^T f(x)$$

for every $y \in C$,

so x solves VI(f,C), where the first inequality holds because $\nabla f(x)$ is positive definite.

Auslender shows that under the conditions on f and C stated above, the descent algorithm with $D(x) = y_x-x$ and $T_k(x,d) = \lambda_k$ as described in (2) above can be used to find a stationary point and, hence, a solution to VI(f,C). The algorithm can be stated as follows:

Algorithm

Let $\{\lambda_k\}$ be a sequence of positive real numbers satisfying $\Sigma\lambda_k = +\infty$ and $\lim_{k \to \infty} \lambda_k = 0$.

Step 0: Select $x^0 \in C$. Set k=0.

Step 1: Given x^k, let y^k be a solution to the minimization problem

$$\min_{y \in C} (y-x^k)^T f(x^k).$$

If $(y^k)^T f(x^k) = (x^k)^T f(x^k)$, stop: x^k is a solution to VI(f,C).

Otherwise, go to Step 2.

Step 2: Given x^k and y^k, let $x^{k+1} + \lambda_k(y^k - x^k)$.

Go to step 1 with k=k+1. ∎

In Chapter 4, we show that this algorithm can be interpreted as a modification of the Frank-Wolfe [1956] algorithm.

CHAPTER 3

THE GENERALIZED STEEPEST DESCENT METHOD FOR UNCONSTRAINED VARIATIONAL INEQUALITY PROBLEMS

3.1 Introduction

In this chapter we consider the unconstrained variational inequality problem $VI(f,R^n)$:

$$\text{find } x^* \varepsilon R^n \text{ satisfying } (x-x^*)^T f(x^*) \geq 0 \text{ for every } x \varepsilon R^n, \tag{3.1}$$

where $f:R^n \to R^n$ is continuously differentiable and uniformly monotone. (If f is affine, uniform monotonicity reduces to strict monotonicity). The unconstrained problem seeks a zero of the mapping f, since $(x-x^*)^T f(x^*) \geq 0$ for every x in R^n if and only if $f(x^*)=0$.

If $\nabla f(x)$ is symmetric for every x in R^n, then $f(x) = \nabla F(x)$ for some uniformly convex functional $F:R^n \to R$ (F is a strictly convex quadratic function when f is a strictly monotone affine mapping), and the unique solution x^* satisfying $f(x^*)=0$ solves the unconstrained convex minimization problem

$$\min_{x \varepsilon R^n} F(x). \tag{3.2}$$

In this case, the solution to the unconstrained variational inequality problem can be found by using the steepest descent method to find the point x^* at which F achieves its minimum over R^n.

For notational convenience in describing this algorithm and related results in this chapter, for any two points x and d in R^n, we let [x;d] denote the ray emanating from x in the direction d; i.e.,

$$[x;d] = \{y : y = x + \theta d,\ \theta \geq 0\}.$$

Steepest Descent Algorithm for Unconstrained Minimization Problems

Step 0: Select $x^o \varepsilon R^n$. Set $k=0$.

Step 1: Compute $-\nabla F(x^k)$.

If $\nabla F(x^k)=0$, then stop: $x^k=x^*$.

Otherwise, go to Step 2.

Step 2: Find $x^{k+1} = \arg\min\{F(x): x \varepsilon [x^k; -\nabla F(x^k)]\}$. Go to Step 1 with $k = k+1$. ∎

Curry [1944] and Courant [1943] have given early expositions on this classical method. Curry attributes the origin of the method to Cauchy [1847], while Courant attributes it to Hadamard [1907]. The following result (see, for example, Polak [1971]) summarizes the algorithm's convergence properties.

Theorem 3.1

Assume that the level set $S(F,x^o)=\{x : x \varepsilon R^n,\ F(x) \leq F(x^o)\}$ is bounded and that $F:R^n \rightarrow R^1$ is continuously differentiable on the convex hull of $S(F,x^o)$. Then either the sequence $\{x^k\}$ constructed by the steepest descent algorithm is finite, terminating at a point x^N satisfying $\nabla F(x^N)=0$, or it is infinite, and every limit point x^* of the sequence $\{x^k\}$ (and there exists at least one) satisfies $\nabla F(x^*)=0$.

Local rate of convergence results can be obtained by approximating $F(x)$ by a quadratic function and using the following result. (See, for example, Bertsekas [1982].)

Theorem 3.2

If $F(x) = (1/2)(x-x^*)^T Q(x-x^*)$ and Q is a positive definite symmetric nxn matrix, then the sequence $\{x^*\}$ generated by the steepest descent algorithm satisifes

$$F(x^{k+1}) \leq \left[\frac{A-a}{A+a}\right]^2 F(x^k), \tag{3.3}$$

where A and a are, respectively, the largest and smallest eigenvalues of Q. Consequently, the sequence $\{x^k\}$ satisfies

$$\|x^{k+1}-x^*\|_Q \leq \frac{A-a}{A+a} \| x^k-x^* \|_Q,$$

$\| x \|_Q = (x^T Qx)^{\frac{1}{2}}$.

When f is a gradient mapping, we can reformulate $VI(f,R^n)$ as an equivalent minimization problem as indicated above and use the steepest descent algorithm to solve the minimization problem; equivalently, we can restate the steepest descent algorithm in a form that can be applied directly to the variational inequality problem. To do so, we eliminate any reference to F(x) in the algorithm and refer only to $f(x) = \nabla F(x)$. In Step 1, we compute $- f(x^k) = -\nabla F(x^k)$, terminating only if $f(x^k) = \nabla F(x^k) = 0$. Since F(x) is convex, x^{k+1} solves the one-dimensional optimization problem in Step 2 if and only if the directional derivative of F at x^{k+1} is nonnegative in all feasible directions. That is,

$$x^{k+1}\varepsilon[x^k;-\nabla F(x^k)] \text{ satisfies } (x-x^{k+1})^T\nabla F(x^{k+1})\geq 0 \text{ for every } x\varepsilon[x^k;-\nabla F(x^k)],$$

or, equivalently,

$x^{k+1}\varepsilon[x^k;-f(x^k)]$ satisfies $(x-x^{k+1})^T f(x^{k+1})\geq 0$ for every $x\varepsilon[x^k;-f(x^k)]$.

With this reformulation, the following "generalized" steepest descent is applicable to any unconstrained variational inequality problem.

Generalized Steepest Descent Algorithm for the Unconstrained Variational Inequality Problem

Step 0: Select $x^o\varepsilon R^n$. Set $k = 0$.

Step 1: Compute $-f(x^k)$.

If $f(x^k) = 0$, stop; $x^k = x^*$.

Otherwise, go to Step 2.

Step 2: Find $x^{k+1}\varepsilon[x^k;-f(x^k)]$ satisfying

$$(x-x^{k+1})^T f(x^{k+1})\geq 0 \text{ for every } x\varepsilon[x^k;-f(x^k)].$$

Go to Step 1 with $k=k+1$. ▮

Our previous observations establish the following result.

Lemma 3.1

When $f(x) = \nabla F(x)$ for every $x\varepsilon R^n$, the steepest descent algorithms for the unconstrained minimization problem and for the unconstrained variational inequality problem are equivalent.

The generalized steepest descent algorithm will not solve every unconstrained variational inequality problem, even if the underlying map is uniformly monotone. If f is not a gradient mapping, the iterates generated by the algorithm can cycle or diverge. The following example illustrates this type of behavior.

Example 3.1

Let $f(x) = Mx$, where $x \varepsilon R^2$ and $M = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

Since M is positive definite, f is uniformly monotone. f is not, however, a gradient mapping, since $\nabla f(x) = M$ is not symmetric. Let $x^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and consider the progress of the generalized steepest descent algorithm. As long as $x^k \neq x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, the one-dimensional variational inequality subproblem on the k^{th} iteration will solve at the point x^{k+1} at which the vector $f(x^{k+1})$ is orthogonal to $-f(x^k)$, the direction of movement. In this example, $-f(x^0) = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$, which implies that $x^1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, since $f\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ is orthogonal to $\begin{pmatrix} 0 \\ -2 \end{pmatrix}$. Similarly, $x^2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ because $f\begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$, $x^3 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ because $f\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$, and $x^4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = x^0$ because $f\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$. Thus, in this case, the algorithm cycles about the four points $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

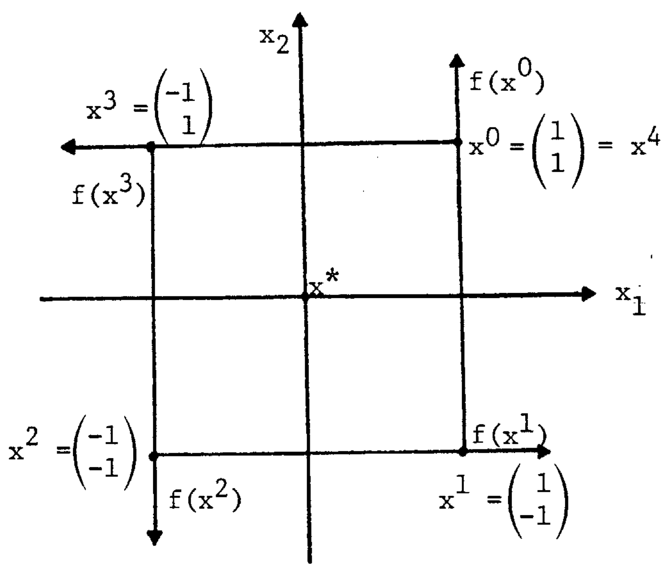


Figure 3.1

The Steepest Descent Iterates Need Not Converge If M is Asymmetric

The iterates produced by the generalized steepest descent algorithm do not converge in this example because the matrix M is "too asymmetric": the off-diagonal entries are too large in absolute value in comparison to the diagonal entries. Define the matrix M_p as $M_p = \begin{bmatrix} 1 & -p \\ p & 1 \end{bmatrix}$. If p=0, the generalized steepest descent method will converge (since $\nabla f(x) = M_0 = I$ is symmetric); if p=1, the above example shows that it will not converge. (Note that M_p is positive definite for all values of p, since $\hat{M}_p = I$ for every p.) An analysis of the vector field of f for various values of p suggests that the iterates move closer to the solution $x^*=0$ if and only if $|p|<1$. In fact, if $f(x) = M_p x$, then $x^{k+1} = p\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} x^k$. Therefore, the iterates converge to the solution if and only if $|p|<1$.

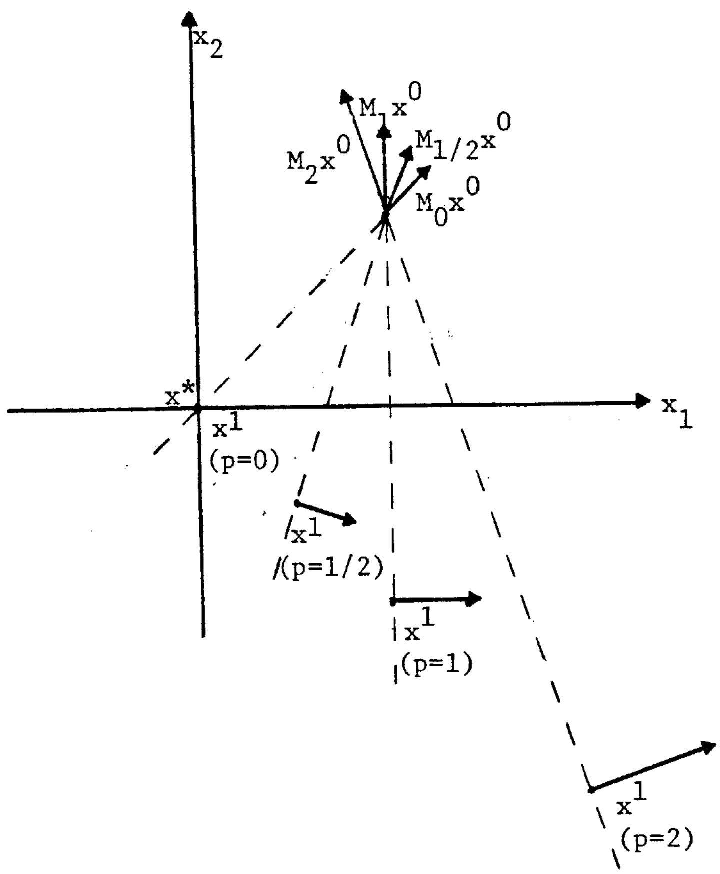


Figure 3.2

A Comparison of Movement Directions for Different Values of p.

In the following analysis, we investigate conditions on f that ensure that the generalized steepest descent algorithm solves an unconstrained variational inequality problem, even when it cannot be reformulated as an equivalent minimization problem. In Section 3.2 we consider a simplified problem setting in which f is an affine, strictly monotone mapping. Section 3.3 extends these results by assuming that f is a nonlinear, uniformly monotone mapping. Section 3.4 shows that the convergence conditions for the generalized steepest descent method may be weakened considerably if the problem mapping is scaled in an appropriate manner. Finally, Section 3.5 extends the results for the generalized steepest descent method to more general gradient methods.

In developing these results, we parameterize the ray $[x^k;-f(x^k)]$ as $x^k-\theta f(x^k)$ for $\theta \geq 0$, and refer to the value θ_k satisfying

$$x^{k+1} = x^k - \theta_k f(x^k)$$

as the step length at the $k^{\underline{th}}$ iteration.

3.2 The Generalized Steepest Descent Algorithm for Unconstrained Problems with Affine Maps

In this section, we consider the unconstrained variational inequality problem VI(f, R^n) where f is a strictly monotone affine map. Thus, we let $f(x)=Mx-b$, where M is an nxn real matrix and $b \varepsilon R^n$, and assume that M is positive definite.

3.2.1 Convergence of the Generalized Steepest Descent Method

When f is affine, we can easily find a closed form expression for the step length θ_k on the $k^{\underline{th}}$ iteration.

Lemma 3.2

Assume that $f(x) = Mx-b$. Let x^k be the $k^{\underline{th}}$ iterate generated by the generalized steepest descent method. Then, if x^k does not solve the problem, the steplength determined on the $k^{\underline{th}}$ iteration is

$$\theta_k = \frac{(Mx^k-b)^T(Mx^k-b)}{(Mx^k-b)^T M(Mx^k-b)} . \tag{3.4}$$

Proof

Step 2 of the algorithm determines $x^{k+1} \varepsilon [x^k; -f(x^k)]$ satisfying

$$(x-x^{k+1})^T f(x^{k+1}) \geq 0 \text{ for every } x \varepsilon [x^k; -f(x^k)]. \tag{3.5}$$

If $x^{k+1} = x^k$, then equation (3.5) with $x = x^k - f(x^k)$ indicates that $-f^T(x^k)f(x^k) \geq 0$. But this implies that $f(x^k) = 0$, so the algorithm would have terminated in Step 1. Hence, assume that $x^{k+1} \neq x^k$, and, therefore, $\theta_k \neq 0$.

Substituting $x = x^k - \theta f(x^k)$ and $x^{k+1} = x^k - \theta_k f(x^k)$ into (3.5) gives

$$(\theta_k - \theta)\ f^T(x^k) f(x^k - \theta_k f(x^k)) \geq 0.$$

Since this inequality is valid for all $\theta \geq 0$, it is equivalent to the condition $f^T(x^k)f(x^k-\theta_k f(x^k))=0$. Substituting $f(x)=Mx-b$ into this expression gives

$$\begin{aligned} & f^T(x^k)f(x^k-\theta_k f(x^k)) \\ & =[Mx^k-b]^T[M(x^k-\theta_k(Mx^k-b))-b] \\ & =[Mx^k-b]^T[Mx^k-b]-\theta_k[Mx^k-b]^T M[Mx^k - b] \\ & = 0 . \end{aligned}$$

This last equality shows that θ_k is given by expression (3.4). ∎

When f is a gradient mapping, convergence of the steepest descent algorithm follows from the fact that $F(x^k)$ is a descent function for the algorithm, where $\nabla F(x) = f(x)$. When the Jacobian of f(x) is not symmetric, however, no function F(x) satisfies $\nabla F(x) = f(x)$, so, in general, this proof of convergence does not apply. Instead, we will establish convergence of the generalized steepest descent method by showing that the iterates produced by the algorithm contract toward the solution with respect to the $\hat{M}$ norm, where $\hat{M} = (1/2)(M+M^T)$ denotes the symmetric part of the matrix M. The $\hat{M}$ norm is a natural choice for establishing convergence because it corresponds directly to the descent function F(x) in the case where f(x) = Mx-b and M is symmetric. In this case, $F(x) = (1/2)x^TMx-b^Tx$, while $\|x-x^*\|_M^2 = (x-x^*)^TM(x-x^*) = 2F(x) + x^{*T}Mx^*$; i.e., F(x) is a descent function for the algorithm if and only if $\|x-x^*\|_{\hat{M}}$ is a descent function.

The following theorem states necessary and sufficient conditions on the matrix M for the steepest descent method to contract with respect to the $\hat{M}$ norm.

<u>Theorem 3.3</u>

Let M be a positive definite matrix, and f(x) = Mx-b. Then the iterates produced by the generalized steepest descent method is guaranteed to contract in $\hat{M}$ norm towards the solution x^* to the problem $VI(f,R^n)$ if and only if the matrix M^2 is positive definite.

Furthermore, the contraction constant is given by

$$r = \left[1 - \inf_{x\neq 0}\left(\frac{(Mx)^T(Mx)}{x^TMx}\,\frac{x^TMMx}{(Mx)^TM(Mx)}\right)\right]^{1/2}. \tag{3.6}$$

Proof:

For ease of notation, let $x \neq x^*$ be the $k^{\underline{th}}$ iterate produced by the algorithm, let $\theta = \theta_k$, and let $\bar{x}$ be the $(k+1)^{st}$ iterate, i.e., $\bar{x} = x - \theta(Mx - b)$, where $\theta = \dfrac{(Mx-b)^T(Mx-b)}{(Mx-b)^T M(Mx-b)}$. We show below that there exists a real number $r \varepsilon [0,1)$ that is independent of x and satisfies $\| \bar{x} - x^* \|_{\hat{M}} \leq r \; \| x - x^* \|_{\hat{M}}$. Because r must satisfy

$$r \geq T(x) := \frac{\| \bar{x}-x^* \|_{\hat{M}}}{\| x-x^* \|_{\hat{M}}}$$

for every $x \neq x^*$, we define

$$r := \sup_{x \neq x^*} T(x).$$

r is clearly nonnegative, since $T(x) > 0$ for every $x \neq x^*$.

We now show that $r<1$. Because $\| z \|^2_{\hat{M}} = z^T Mz$ for every z, we have that $\| x-x^* \|_{\hat{M}} = [(x-x^*)^T M(x-x^*)]^{\frac{1}{2}}$, and

$$\| \bar{x}-x^* \|_{\hat{M}} = [(x-\theta(Mx-b)-x^*)^T M(x-\theta(Mx-b)-x^*)]^{\frac{1}{2}}$$

$$= [(x-x^*)^T M(x-x^*) - \theta(Mx-b)^T M(x-x^*) - \theta(x-x^*)^T M(Mx-b) + \theta^2(Mx-b)^T M(Mx-b).$$

$$\text{Thus, } T(x) = \left[1 - \frac{\theta(Mx-b)^T M(x-x^*) + \theta(x-x^*)^T M(Mx-b) - \theta^2(Mx-b)^T M(Mx-b)}{(x-x^*)^T M(x-x^*)} \right]^{\frac{1}{2}}.$$

Because the solution to the problem is $x^* = M^{-1}b$,

$$Mx-b = M(x-x^*) \text{ and, hence, } \theta = \frac{[M(x-x^*)]^T[M(x-x^*)]}{[M(x-x^*)]^T M[M(x-x^*)]}.$$

Substituting this value for θ in the previous expression for T(x), we see that the first and third terms in the numerator of the fraction cancel.

Consequently,

$$T(x) = \left[1 - \frac{[M(x-x^*)]^T[M(x-x^*)] \cdot [(x-x^*)^T M^2 (x-x^*)]}{[M(x-x^*)]^T M[M(x-x^*)] \cdot [(x-x^*)^T M(x-x^*)]}\right]^{\frac{1}{2}}$$

$$= [1-R(y)]^{\frac{1}{2}},$$

where $y = x-x^* \neq 0$ and $R(y) := \dfrac{[(My)^T(My)][y^TM^2y]}{[(My)^TM(My)][y^TMy]}$.

Therefore, $r = \sup\limits_{x \neq x^*} T(x) = \sup\limits_{y\neq 0} [1-R(y)]^{\frac{1}{2}}$.

To complete the proof of the theorem, we show that $r<1$ if and only if M^2 is positive definite. We first note that $r = [1-\inf\limits_{y\neq 0} R(y)]^{\frac{1}{2}}<1$ if and only if $\inf\limits_{y\neq 0} R(y)>0$.

Now if M^2 is positive definite, then $\widehat{M^2}$ is positive definite, so

$$\inf_{y\neq 0} R(y) = \inf_{y\neq 0} \left\{ \frac{\dfrac{y^TM^2y}{y^Ty}}{\dfrac{y^TMy}{y^Ty} \cdot \dfrac{(My)^TM(My)}{(My)^T(My)}} \right\}$$

$$\geq \frac{\inf\limits_{y\neq 0}\left\{\dfrac{y^TM^2y}{y^Ty}\right\}}{\sup\limits_{y\neq 0}\left\{\dfrac{y^TMy}{y^Ty}\right\} \cdot \sup\limits_{y\neq 0}\left\{\dfrac{(My)^TM(My)}{(My)^T(My)}\right\}}$$

$$= \frac{\lambda_{min}(\widehat{M^2})}{[\lambda_{max}(\hat{M})]^2} = \frac{\lambda_{min}(\widehat{M^2})}{\lambda_{max}(\hat{M}^2)} \quad ,$$

where $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote, respectively, the minimum and maximum eigenvalues of the real symmetric matrix A.

Since $\widehat{M^2}$ is positive definite, $\widehat{M^2}$ is positive definite and symmetric and, hence, has real and positive eigenvalues. Similarly, M positive definite ensures that the eigenvalues of $\hat{M}$ are real and positive. Consequently, $\inf_{y \neq 0} R(y) > 0$, and, hence $r < 1$.

Conversely, if M^2 is not positive definite, then $y^T M^2 y \leq 0$ for some vector $y \neq 0$. Because M is positive definite, $(My)^T M(My) > 0$ and $y^T My > 0$. Moreover, $y \neq 0$ ensures that $(My)^T(My) > 0$, and, therefore, $R(y) \leq 0$, which implies that $r \geq 1$. ∎

Corollary

The contraction constant r is bounded from above by

$$\bar{r} = \left[1 - \frac{\lambda_{min}(\hat{M}^{-1}\widehat{M^2})}{\lambda_{max}\hat{M}}\right]^{\frac{1}{2}} = \left[1 - \frac{\lambda_{min}[(\hat{M}^{\frac{1}{2}})^{-T}\widehat{M^2}(\hat{M}^{\frac{1}{2}})^{-1}]}{\lambda_{max}\hat{M}}\right]^{\frac{1}{2}} . \tag{3.7}$$

Proof

r is defined by $r = [1 - \inf_{y \neq 0} R(y)]^{\frac{1}{2}}$, where

$$\inf_{y \neq 0} R(y) = \inf_{y \neq 0} \left\{ \frac{y^T M^2 y \cdot (My)^T(My)}{y^T My \cdot (My)^T M(My)} \right\} \geq \frac{\inf_{y \neq 0}\left\{\frac{y^T M^2 y}{y^T My}\right\}}{\sup_{y \neq 0}\left\{\frac{(My)^T M(My)}{(My)^T My}\right\}} . \tag{3.8}$$

$$\begin{aligned}
\inf_{y \neq 0} \left\{\frac{y^T M^2 y}{y^T My}\right\} &= \inf_{y \neq 0}\left\{\frac{y^T \widehat{M^2} y}{y^T \hat{M} y}\right\} \\
&= \inf_{y \neq 0}\left\{\frac{y^T \widehat{M^2} y}{y^T (\hat{M}^{\frac{1}{2}})^T \hat{M}^{\frac{1}{2}} y}\right\} \\
&= \inf_{y \neq 0}\left\{\frac{y^T \widehat{M^2} y}{(\hat{M}^{\frac{1}{2}} y)^T (\hat{M}^{\frac{1}{2}} y)}\right\} \\
&= \inf_{z \neq 0}\left\{\frac{z^T (\hat{M}^{\frac{1}{2}})^{-T} \widehat{M^2} (\hat{M}^{\frac{1}{2}})^{-1} z}{z^T z}\right\}
\end{aligned}$$

$$= \lambda_{min}(A), \text{ where } A = (\hat{M}^{\frac{1}{2}})^{-T}\widehat{M^2}(\hat{M}^{\frac{1}{2}})^{-1}.$$

Now λ is a an eigenvalue of A if and only if λ is an eigenvalue of $\hat{M}^{-1}\widehat{M^2}$, since

$$\begin{aligned}
&\det(A-\lambda I) = 0 \\
&\leftrightarrow \det(\hat{M}^{\frac{1}{2}})^T(A-\lambda I)(\hat{M}^{\frac{1}{2}}) = 0 \\
&\leftrightarrow \det(\widehat{M^2} - \hat{M}\lambda) = 0 \\
&\leftrightarrow \det(\hat{M}^{-1}(\widehat{M^2}-\hat{M}\lambda)) = 0 \\
&\leftrightarrow \det(M^{-1}\widehat{M^2}-\lambda I) = 0 .
\end{aligned}$$

Thus, $$\inf_{y\neq 0}\left\{\frac{y^TM^2y}{y^TMy}\right\} = \lambda_{min}(A) = \lambda_{min}(\hat{M}^{-1}\widehat{M^2}). \tag{3.9}$$

Finally, since $z^TMz = z^T\hat{M}z$ for any vector z, and $z = My$ is nonzero if and only if y is nonzero,

$$\sup_{y\neq 0}\left\{\frac{(My)^TM(My)}{(My)^TMy}\right\} = \lambda_{max}(\hat{M}). \tag{3.10}$$

The result follows from (3.8), (3.9) and (3.10). ∎

3.2.2 Discussion of M^2

The theorem indicates that the key to convergence of the generalized steepest descent method is the matrix M^2. If the positive definite matrix M is symmetric, the convergence of the steepest descent algorithm for unconstrained convex minimization problems follows immediately: $M^2=M^T \cdot M$ is positive definite because M, being positive definite, is nonsingular. In general, the condition that the square of the positive definite matrix M be positive definite imposes a restriction on the degree to which M can differ from M^T. To see this, first note that M^2 is positive definite if and only if

$$x^T M^2 x = (M^T x)^T (Mx) > 0 \text{ for every } x \neq 0.$$

Thus, M^2 is positive definite if and only if for every nonzero vector x, the angle between the vectors $(M^T x)$ and (Mx) is less than 90°.

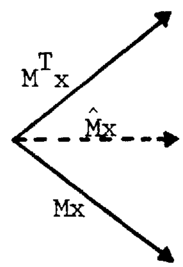


Figure 3.3

M^2 is Positive Definite If and Only If $(M^T x)^T Mx>0$ for Every Nonzero x.

The positive definiteness of M^2 does not imply an absolute upper bound on the quantity $\| M-M^T \|$ for any norm $\| \cdot \|$, because we can always increase this quantity by multiplying M by a constant. If M^2 is positive definite, then the normalized quantity $\| M-M^T \| / \| M+M^T \|$ must be less than 1. This result follows from the following proposition.

Proposition 3.1 (Anstreicher [1984])

Let M be an nxn real matrix. Then, for any norm $\| \cdot \|$, M^2 is positive definite if and only if $\| (M-M^T)x \| < \| (M+M^T)x \|$ for every $x \neq 0$.

Proof

$4M^2 = (M+M^T)^2 + (M-M^T)^2 + (M-M^T)(M+M^T) + (M+M^T)(M-M^T)$.

But, $[(M-M^T)(M+M^T)] = -(M+M^T)(M-M^T)$.

Hence, $M^2 + (M^2)^T = \frac{1}{2}[(M+M^T)^2 + (M-M^T)^2]$.

$$\begin{aligned}
\text{Thus, } x^T(M^2)x>0 \quad &\leftrightarrow \quad x^T(M^2 + (M^2)^T)x>0 \\
&\leftrightarrow \quad x\ (M+M^T)^2x + x^T(M-M^T)^2x>0 \\
&\leftrightarrow \quad x^T(M+M^T)^T(M+M^T)x > x^T(M-M^T)^T(M-M^T)x \\
&\leftrightarrow \quad \|(M+M^T)x\| > \|(M-M^T)x\| .
\end{aligned}$$

∎

In particular, if M^2 is positive definite, then

$$\begin{aligned}
\|M-M^T\| &= \max_{\|x\|=1} \|(M-M^T)x\| = \|(M-M^T)\bar{x}\| < \|(M+M^T)\bar{x}\| \\
&\le \max_{\|x\|=1} \|(M+M^T)x\| = \|M+M^T\|, \text{ where } \bar{x} = \operatorname*{argmax}_{\|x\|=1} \|(M-M^T)x\|.
\end{aligned}$$

Consequently, $\|M-M^T\| / \|M+M^T\| < 1$ for any norm.

3.2.3 Discussion of the Bound on the Contraction Constant

Let us return to the problem defined in Example 3.1. The mapping $f(x) = M_p x$ is affine and strictly monotone, since $M_p = \begin{bmatrix} 1 & -p \\ p & 1 \end{bmatrix}$ is positive definite. For this example, $M_p^2 = \begin{bmatrix} 1-p^2 & -2p \\ 2p & 1-p^2 \end{bmatrix}$, and $M_p^2 = \begin{bmatrix} 1-p^2 & 0 \\ 0 & 1-p^2 \end{bmatrix}_2 = (1-p^2)I$. Thus, as we would expect from the observations after the example, M_p^2 is positive definite if and only if $|p| < 1$. Moreover,

$$\frac{\|M_p - M_p^T\|_2}{\|M_p + M_p^T\|_2} = \frac{\left\| \begin{matrix} 0 & -2p \\ 2p & 0 \end{matrix} \right\|_2}{\left\| \begin{matrix} 2 & 0 \\ 0 & 2 \end{matrix} \right\|_2} = |p| < 1 \text{ if and only if } M_p^2 \text{ is positive}$$

definite.

For this example, the upper bound on the contraction constant given in the corollary is tight. To see this, first note that $\hat{M}_p = I$, so the $\hat{M}_p$ norm is equivalent to the Euclidean norm. Recall from the example that $x^{k+1} = p\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}x^k$ and $x^* = 0$. Thus,

$$\begin{aligned}
\| x^{k+1} - x^* \|_{\hat{M}} &= \| x^{k+1} \|_2 \\
&= \left(p^2 (x^k)^T \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}^T \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} x^k \right)^{\frac{1}{2}} \\
&= (p^2 (x^k)^T I \; x^k)^{\frac{1}{2}} \\
&= |p| \; \| x^k \|_2 \\
&= |p| \; \| x^k - x^* \|_{\hat{M}} ,
\end{aligned}$$

and the contraction constant for the problem is $|p|$. The bound given by the corollary is also $|p|$, because $\lambda_{min}(\hat{M}^{-1}\hat{M}^2) = \lambda_{min} \begin{bmatrix} 1-p^2 & 0 \\ 0 & 1-p^2 \end{bmatrix} = 1 - p^2$ and $\lambda_{max}(\hat{M}) = 1$ give $\bar{r} = [1-(1-p^2)]^{\frac{1}{2}} = |p|$.

For affine problems with symmetric matrices, the bound $\bar{r}$ on the contraction constant r may be quite loose. If M is symmetric, a tighter upper bound on r found by diagonalizing M and applying the Kantorovich inequality [see, for example, Luenberger (1973)] is

$r_s = \dfrac{\lambda_{max}(M) - \lambda_{min}(M)}{\lambda_{max}(M) + \lambda_{min}(M)}$. In terms of the condition number $k = \lambda_{max}(M)/\lambda_{min}(M)$, $r_s = (k-1)/(k+1)$, while $\bar{r} = [(k-1)/k]^{\frac{1}{2}}$. Thus, for example, if $k=1$, then $r_s = \bar{r} = 0$; if $k = 1.5$, then $r_s = 0.2$ and $\bar{r} = 0.58$; if $k = 3$, then $r_s = 0.5$ and $\bar{r} = 0.82$; and if $k = 10$, then $r_s = 0.82$ and $\bar{r} = 0.95$. This tighter upper bound on r cannot be derived in the same way if M is not symmetric. A matrix M can be decomposed into its spectral decomposition if and only if M is normal, which is true for a real matrix M if and only if $M^TM = MM^T$. Thus, if M is not symmetric, we cannot necessarily diagonalize M. If M is not unitarily equivalent to a diagonal matrix, then we cannot use the Kantorovich inequality to obtain the upper bound r_s.

3.2.4 Sufficient Conditions for M^2 to be Positive Definite

We now seek easy to verify conditions on the matrix M that will ensure that the matrix M^2 is positive definite. The following example shows that the double diagonal dominance condition, a necessary and sufficient condition for convergence for the problem in Example 3.1, is not in general a sufficiently strong condition for M^2 to be positive definite.

Let

$$M = \begin{bmatrix} 2 & 0 & 0 \\ .99 & 1 & 0 \\ .99 & 0 & 1 \end{bmatrix}. \text{ Then } M^2 = \begin{bmatrix} 4 & 0 & 0 \\ 2.97 & 1 & 0 \\ 2.97 & 0 & 1 \end{bmatrix} \text{ and } 2\widehat{M^2} = \begin{bmatrix} 8 & 2.97 & 2.97 \\ 2.97 & 2 & 0 \\ 2.97 & 0 & 2 \end{bmatrix}.$$

Since $\det(2\widehat{M^2}) = -3.2836$, $\widehat{M^2}$ is not positive definite, and, therefore, M^2 is not positive definite.

The norm condition $\| D^{-\frac{1}{2}} B D^{-\frac{1}{2}} \|_2 < 1$, where $D = \text{diag}(M)$ and $B=M-D$, is a sufficient condition to ensure that the linear Jacobi method will solve an unconstrained variational inequality problem with an affine map.

($\| D^{-\frac{1}{2}}BD^{-\frac{1}{2}}\|_2 < 1$ implies the usual condition for convergence of the Jacobi method for linear equations, $\rho(D^{-1}B)<1$, where $\rho(D^{-1}B)$ is the spectral radius of $D^{-1}B$, because $\rho(D^{-1}B) \leq \| D^{-1}B\|_D = \| D^{-\frac{1}{2}}BD^{-\frac{1}{2}}\|_2$.) Pang and Chan [1981] have shown that if M is doubly diagonally dominant, then $\| D^{-\frac{1}{2}}BD^{-\frac{1}{2}}\|_2 < 1$. The above example, therefore, also shows that $\| D^{-\frac{1}{2}}BD^{-\frac{1}{2}}\|_2 < 1$ is not a strong enough condition on M to ensure that M^2 is positive definite.

As shown by the following theorem, stronger double diagonal dominance conditions imposed on M guarantee that M^2 is doubly diagonally dominant, which, in turn, implies that M^2 is positive definite.

Theorem 3.4

Let $M = (M_{ij})$ be an nxn matrix with positive diagonal entries. If for every $i = 1,2,\ldots,n$,

$$\sum_{j\neq i} |M_{ij}| < ct \text{ and } \sum_{j\neq i} |M_{ji}| < ct, \text{ where } t = \frac{\min\{(M_{ii})^2 : i=1,\ldots,n\}}{\max\{M_{ii} \quad : i=1,\ldots,n\}}$$

and $c = \sqrt{2} - 1$,

then M^2 is doubly diagonally dominant, and, therefore, positive definite, and M is doubly diagonally dominant, and, therefore, positive definite.

Proof

Let M_{ij} be the $(i,j)^{th}$ element of M.

The $(i,j)^{th}$ element of M^2 is

$$(M^2)_{ij} = \sum_{k=1}^{n} M_{ik}M_{kj} = \begin{cases} M_{ii}^2 + \sum_{k\neq i} M_{ik}M_{ki} & \text{if } i=j \\ \\ M_{ii}M_{ij} + M_{ij}M_{jj} + \sum_{k\neq i,j} M_{ik}M_{kj} & \text{if } i\neq j. \end{cases}$$

To show that M^2 is doubly diagonally dominant, we must show that

$$(M^2)_{ii} > \sum_{j\neq i} |(M^2)_{ij}| \text{ and } (M^2)_{ii} > \sum_{j\neq 1} |(M^2)_{ji}|,$$

i.e., that
$$M_{ii}^2 > -\sum_{k\neq i} M_{ik}M_{ki} + \sum_{j\neq i} |M_{ii}M_{ij}+M_{ij}M_{jj} + \sum_{k\neq i,j} M_{ik}M_{kj}|, \tag{3.11}$$

and
$$M_{ii}^2 > -\sum_{k\neq i} M_{ik}M_{ki} + \sum_{j\neq i} |M_{ii}M_{ji} + M_{ji}M_{jj} + \sum_{k\neq i,j} M_{jk}M_{ki}|. \tag{3.12}$$

To show that (3.11) holds, it is enough (by the triangle inequality) to show that

$$M_{ii}^2 > -\sum_{k\neq i} M_{ik}M_{ki} + \sum_{j\neq i} [\,|M_{ii}M_{ij}|+|M_{ij}M_{jj}|+ \sum_{k\neq i,j} |M_{ik}M_{kj}|\,],$$

which is true (by Cauchy's Inequality and the triangle inequality) if

$$M_{ii}^2 > \sum_{k\neq i} |M_{ik}|\,|M_{ki}| + M_{ii}\sum_{j\neq i} |M_{ij}| + \sum_{j\neq i} M_{jj}|M_{ij}| + \sum_{j\neq i}\sum_{k\neq i,j} |M_{ik}|\,|M_{kj}|.$$

Because the last term in the righthand side of the above expression is equal to $\sum_{k\neq i}\sum_{j\neq i,k} |M_{ik}|\,|M_{kj}|$, the sum of the first and last terms in the righthand side is

$$\sum_{k\neq i} |M_{ik}| \left[|M_{ki}| + \sum_{j\neq i,k} |M_{kj}| \right] = \sum_{k\neq i} |M_{ik}| \left[\sum_{j\neq k} |M_{kj}| \right].$$

Consequently, to show (3.11) is true, we must show that

$$M_{ii}^2 > \sum_{k\neq i} |M_{ik}| \left[\sum_{j\neq k} |M_{kj}| \right] + M_{ii}\sum_{j\neq i} |M_{ij}| + \sum_{j\neq i} M_{jj}|M_{ij}|. \tag{3.13}$$

To establish (3.13) (and, hence, (3.11)), we introduce a quantity t defined as

$$t = \frac{\min\{M_{ii}^2 : i=1,\ldots,n\}}{\max\{M_{ii} : i=1,\ldots,n\}} .$$

Note that

(i) $t \leq M_{ii}$ for every $i=1,\ldots, n$ (since $t \leq M_{ii}^2/M_{ii} = M_{ii}$ for every i), and

(ii) $t \cdot \mathrm{Max}\{M_{ii}: i=1,\ldots,n\} \leq M_{ii}^2$ for every $i=1,\ldots,n$.

The bounds on the off-diagonal elements of M assumed in the statement of the theorem ensure that the righthand side, RHS, of (3.13) satisfies

$$\begin{aligned} \mathrm{RHS} &< \sum_{k\neq i} |M_{ik}|(ct) + M_{ii}(ct) + \mathrm{Max}\{M_{ii}: i=1,\ldots,n\}(ct), \\ &< c^2t^2 + M_{ii}(ct) + (ct)\ \mathrm{Max}\{M_{ii}: i=1,\ldots,n\} \\ &\leq c^2M_{ii}^2 + cM_{ii}^2 + cM_{ii}^2 \qquad \text{by (i) and (ii).} \end{aligned}$$

Thus, (3.13) holds if $M_{ii}^2(c^2 + 2c) \leq M_{ii}^2$, or, since $M_{ii} > 0$, if $c^2 + 2c-1\leq 0$, which holds if and only if $c\varepsilon[-1-\sqrt{2},\sqrt{2}-1]$. Thus, if $c=\sqrt{2}-1$, (3.13) and, therefore, (3.11) must hold. Similarly, if $c = \sqrt{2} -1$, then (3.12) must hold. These two results establish that M^2 is doubly diagonally dominant.

The double diagonal dominance of M^2 ensures that $\widehat{M^2}$ is doubly diagonally dominant. Because $\widehat{M^2}$ is symmetric and row diagonally dominant, by the Gershgorin Circle Theorem (Gershgorin [1931]), $\widehat{M^2}$ has real, positive eigenvalues. Since $\widehat{M^2}$ is symmetric and has positive eigenvalues, $\widehat{M^2}$ is positive definite, and, therefore, M^2 is positive definite.

The conditions that the theorem imposes on the off-diagonal elements of M also ensure that M is doubly diagonally dominant. Therefore, M is positive definite. ∎

3.2.5 Comparison of Conditions on M

Because the assumption that M is doubly diagonally dominant is stronger than the assumption that M^2 is positive definite, the conditions imposed on M in Theorem 3.4 are likely to be stronger than necessary to show that M^2 is positive definite. In Examples 3.2 and 3.3, we compare the following three conditions:

(1) the conditions of Theorem 3.4;

(2) necessary and sufficient conditions for the matrix M^2 to be doubly diagonally dominant; and

(3) necessary and sufficient conditions for the matrix M^2 to be positive definite.

From the proof of Theorem 3.4, we know that conditions (1) imply conditions (2), and that conditions (2) imply conditions (3). The examples suggest that there is a much larger "gap" between conditions (2) and (3) than between conditions (1) and (2). Thus, it seems that we cannot find conditions much less restrictive than the conditions of the theorem as long as we look for conditions that imply that M^2 is doubly diagonally dominant instead of showing directly that M^2 is positive definite.

Example 3.2

Let $M = \begin{bmatrix} N & 0 & 0 \\ a & 1 & 0 \\ b & 0 & 1 \end{bmatrix}$, where $N>0$.

Then $M^2 = \begin{bmatrix} N^2 & 0 & 0 \\ (N+1)a & 1 & 0 \\ (N+1)b & 0 & 1 \end{bmatrix}$ and $M^2 = \begin{bmatrix} N^2 & \frac{1}{2}(N+1)a & \frac{1}{2}(N+1)b \\ \frac{1}{2}(N+1)a & 1 & 0 \\ \frac{1}{2}(N+1)b & 0 & 1 \end{bmatrix}$.

Condition (1) requires that for $i=1,2,3$, $\sum_{j\neq i} |M_{ij}|$ and $\sum_{j\neq i} |M_{ji}|$ are bounded from above by

$$(\sqrt{2}-1)\,\frac{\text{Min}\{M_{ii}^2:\ i=1,2,3\}}{\text{Max}\{M_{ii}:\ i=1,2,3\}} = \begin{cases} (\sqrt{2}-1)N^2 & \text{if } 0 < N \leq 1 \\ (\sqrt{2}-1)\frac{1}{N} & \text{if } N > 1; \end{cases}$$

i.e.,

$$\text{that } |a| + |b| < \begin{cases} (\sqrt{2}-1)N^2 & \text{if } 0<N\leq 1 \\ (\sqrt{2}-1)\frac{1}{N} & \text{if } N>1 \end{cases} .$$

(The conditions $|a| < 1$, $|b| < 1$ are redundant for all values of N).

Condition (2) requires that

$$|a| < \frac{1}{N+1}\,,\quad |b| < \frac{1}{N+1} \text{ and } |a| + |b| < \frac{N^2}{N+1}\,.$$

(For $N\geq\sqrt{2}$, the third condition is redundant, while for $N\leq 1$, the first two conditions are redundant).

Condition (3) requires (by requiring that the determinant of each positive principle minor of M^2 be positive) that

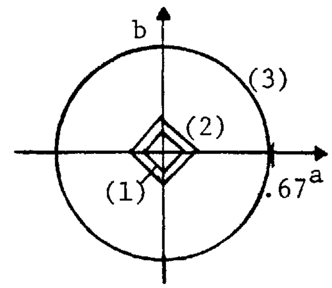
$$a^2 + b^2 < \left(\frac{2N}{N+1}\right)^2 .$$

We now consider the regions in the a-b plane allowed by the restrictions imposed by each of the three conditions for several different values of N.

For $N = \frac{1}{2}$, Condition (1) $\rightarrow |a| + |b| < 0.10$

Condition (2) $\rightarrow |a| + |b| < 0.16$

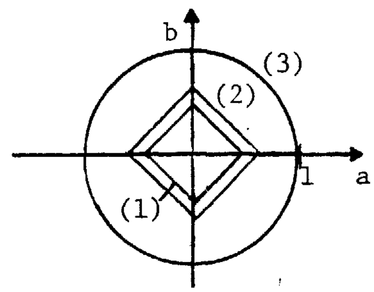
Condition (3) $\rightarrow a^2 + b^2 < (0.67)^2$



For $N = 1$, Condition (1) $\rightarrow |a| + |b| < 0.41$

Condition (2) $\rightarrow |a| + |b| < 0.5$

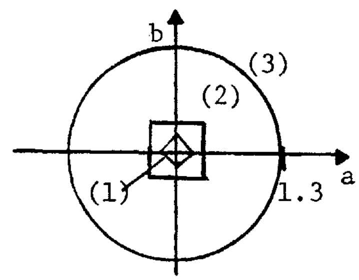
Condition (3) $\rightarrow a^2 + b^2 < 1$



For $N = 2$, Condition (1) $\rightarrow |a| + |b| < 0.21$

Condition (2) $\rightarrow |a| < 0.33$, $|b| < 0.33$

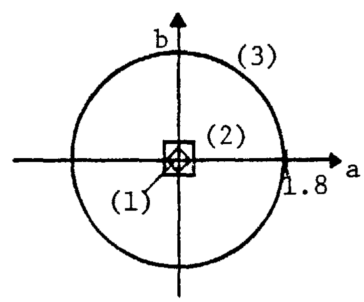
Condition (3) $\rightarrow a^2 + b^2 < (1.33)^2$



For $N = 10$, Condition (1) $\rightarrow |a| + |b| < 0.04$

Condition (2) $\rightarrow |a| < 0.09$, $|b| < 0.09$

Condition (3) $\rightarrow a^2 + b^2 < (1.8)^2$



For this example, as N increases, the separation between conditions (2) and (3) increases, while the disparity between (1) and (2) stays about the same. In particular, as $N \rightarrow \infty$, conditions (1) and (2) drive $|a|$ and $|b|$ to zero. while condition (3) tends to $a^2 + b^2 < 4$. ▮

Example 3.3

Let $M = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & 0 & N \end{bmatrix}$, where $N>0$. Then $M^2 = \begin{bmatrix} 1 & 0 & 0 \\ 2a & 1 & 1 \\ (N+1)b & 0 & N^2 \end{bmatrix}$, and $$\hat{M}^2 = \begin{bmatrix} 1 & a & (\frac{N+1}{2})b \\ a & 1 & 0 \\ (\frac{N+1}{2})b & 0 & N^2 \end{bmatrix}.$$

Condition (1) requires that $|a| + |b| < \begin{cases} (\sqrt{2} - 1)N^2 & \text{if} \quad 0 < N \leq 1 \\ (\sqrt{2} - 1)\frac{1}{N} & \text{if} \quad N > 1 \end{cases}$.

Condition (2) requires that $|a| < \frac{1}{2}$, $|b| < \frac{N^2}{N+1}$, and $2|a| + (N+1)|b| < 1$,

(Note that the second constraint is redundant if $N \geq 1$, while the first is always redundant.)

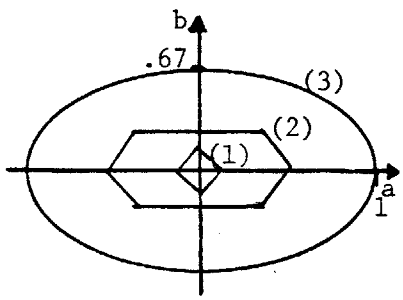
Condition (3) requires that $a^2 + \left(\frac{N+1}{2N}\right)^2 b^2 < 1$.

Again, we consider the regions allowed by these three sets of restrictions for different values of N.

For $N = \frac{1}{2}$, Condition (1) $\rightarrow$ $|a| + |b| < 0.10$

Condition (2) $\rightarrow$ $|b| < 0.17$, $2|a| + 1.5|b| < 1$

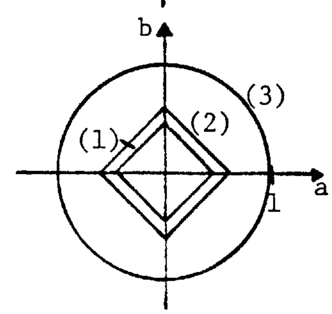
Condition (3) $\rightarrow$ $a^2 + (1.5b)^2 < 1$



For $N = 1$, Condition (1) $\rightarrow$ $|a| + |b| < 0.41$

Condition (2) $\rightarrow$ $|a| + |b| < 0.5$

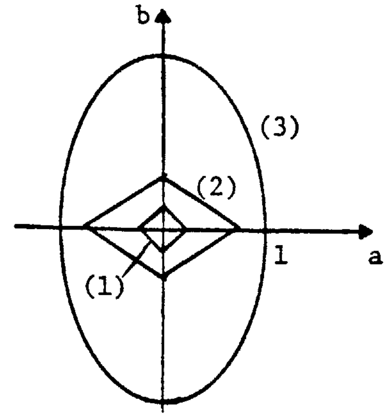
Condition (3) $\rightarrow$ $a^2 + b^2 < 1$



For N = 2, Condition (1) $\rightarrow |a| + |b| < 0.21$

Condition (2) $\rightarrow 2|a| + 3|b| < 1$

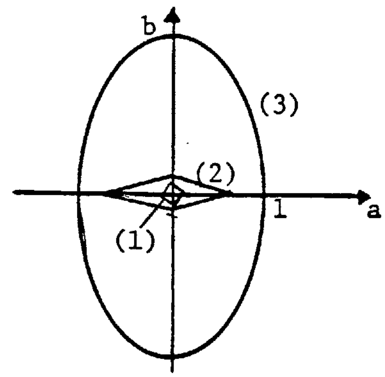
Condition (3) $\rightarrow a^2 + (\frac{3}{4})b)^2 < 1$



For N = 10, Condition (1) $\rightarrow |a| + |b| < 0.04$

(2) $\rightarrow 2|a| + 11\ |b| < 1$

(3) $\rightarrow a^2 + (\frac{11}{20}\, b)^2 < 1$



Again, the separation between conditions (2) and (3) increases considerably as N increases. As $N \rightarrow \infty$, condition (1) drives $|a|$ and $|b|$ to zero, condition (2) tends to $|a| \leq \frac{1}{2}$, $|b| = 0$, and condition (3) tends to $a^2 + (\frac{b}{2})^2 > 1$. In this instance, the discrepancy between conditions (1) and (2) also becomes considerable as $N \rightarrow \infty$. $\blacksquare$

The conditions that Theorem 3.4 imposes on the off-diagonal elements of M are the least restrictive when the diagonal elements of M are all equal. In section 3.4, we show that by scaling the rows or the columns of M so that the scaled matrix has equal diagonal entries, we may be able to weaken considerably the conditions imposed on M.

We close this section with a lemma that shows that the positive definiteness of the matrices M and M^2 is preserved under unitary transformations. A consequence of this result and Theorem 3.3 is that, if M and M^2 are positive definite, then the generalized steepest descent method will solve any unconstrained variational inequality problem defined by a mapping $f(x) = \bar{M}x-b$, where $\bar{M}$ is unitarily equivalent to M. In particular, if $\bar{M}$ is unitarily equivalent to any matrix M that satisfies the conditions of Theorem 3.4, then the generalized steepest descent method will solve VI(f,C), where $f(x) = \bar{M}x-b$.

Lemma 3.3

Let M be a real nxn matrix, and let $\bar{M} = U^TMU$, where U is a real, unitary matrix. Then $\bar{M}$ is positive definite if and only if M is positive definite, and $\bar{M}^2$ is positive definite if and only if M^2 is positive definite.

Proof

Since U is unitary, $U^TU = I$. Note also that since $U^{-1} = U^T$ exists, U is nonsingular. Now M is positive definite if and only if $x^TMx>0$ for every $x\neq 0$, and, hence, if and only if $(Ux)^TM(Ux)>0$ for every $(Ux) \neq 0$. Because U is nonsingular, $Ux \neq 0$ if and only if $x \neq 0$. Thus M is positive definite if and only if $(Ux)^TM(Ux) = x^TU^TMUx = x^T\bar{M}x>0$ for every $x \neq 0$, i.e., if and only if $\bar{M}$ is positive definite.

The second result follows from the first because $\bar{M}^2$ is unitarily equivalent to M^2, which is true because $\bar{M}^2 = (U^T MU)(U^T MU) = U^T M^2 U$. ∎

3.3 The Generalized Steepest Descent Algorithm for Unconstrained Problems with Nonlinear Maps

If $f: R^n \to R^n$ is not affine, strict monotonicity is not a sufficiently strong condition to ensure that a solution to the unconstrained problem $VI(f,R^n)$ exists. If, for example, $n = 1$ and $f(x) = e^x$, then $VI(f,R^1)$ has no solution. Because the ground set R^n over which the problem is formulated is not compact, some type of coercivity condition must be imposed on the mapping f to ensure the existence of a solution. (See, for example, Theorems 1.2 and 1.3). If f is uniformly monotone, then f is strongly coercive: Theorem 1.4 demonstrates that $VI(f,R^n)$ has a solution if f is uniformly monotone and hemicontinuous. Therefore, in this section, we restrict our attention to problems defined by uniformly monotone mappings.

The following theorem establishes conditions under which the generalized steepest descent method will solve an unconstrained variational inequality problem with a nonlinear mapping f. In this case, the key to convergence is the definiteness of the square of the Jacobian of f evaluated at the solution x^*.

Theorem 3.5

Let $f: R^n \to R^n$ be uniformly monotone and twice Gateaux-differentiable. Let $M = \nabla f(x^*)$, where x^* is the unique solution to $VI(f,R^n)$, and assume that M^2 is positive definite. Then, if the initial iterate is sufficiently close in $\hat{M}$ norm to the solution x^*, the sequence of iterates produced by the generalized steepest descent algorithm contracts to the solution in $\hat{M}$ norm.

Proof

To simplify notation, we let $\|\cdot\|$ denote the $\hat{M}$ norm throughout this proof. Let x be the initial iterate. We will show that if the positive real number $\varepsilon := \|x-x^*\|$ is sufficiently small, then the iterates generated by the algorithm contract toward the solution x^*. We assume that $0<\varepsilon<1$.

By Step 2 of the algorithm, the first iterate generated by the algorithm, $\bar{x}$, solves the one-dimensional variational inequality problem on $[x; -f(x)]$, the ray emanating from x in the direction $-f(x)$. The proof of Lemma 3.2 demonstrates that the solution $\bar{x}$ to this one-dimensional problem satisfies $f^T(x)f(\bar{x}) = 0$. Thus, if we parameterize the ray $[x; -f(x)]$ as $x - \theta f(x)$, then $\bar{x} = x - \bar{\theta} f(x)$, where the steplength $\bar{\theta}$ is defined by

$$f^T(x)\, f(x - \bar{\theta}\, f(x)) = 0. \tag{3.14}$$

By Lemma 3.4, which follows, the value $\bar{\theta}$ satisfying 3.14 is unique.

To show that the iterates generated by the algorithm contract towards the solution in $\hat{M}$ norm, we show that there exists a real number $r \varepsilon [0,1)$ that is independent of x and satisfies

$$\|\bar{x} - x^*\| \leq r\, \|x - x^*\|.$$

Because r must satisfy $r \geq T(x) := \dfrac{\|\bar{x} - x^*\|}{\|x - x^*\|}$ for every $x \varepsilon S_\varepsilon = \{x : \|x - x^*\| = \varepsilon\}$, we define $r = \sup_{x \varepsilon S_\varepsilon} T(x)$. r is clearly nonnegative, since $T(x) \geq 0$ for every $x \varepsilon S_\varepsilon$.

We now show that $r<1$. Since we are using the $\hat{M}$ norm,

$$\| x - x^* \|^2 = (x - x^*)^T M(x - x^*), \text{ and}$$

$$\begin{aligned} \| \bar{x} - x^* \|^2 &= (x - \bar{\theta}\, f(x) - x^*)^T M(x - \bar{\theta}\, f(x) - x^*) \\ &= (x - x^*)^T M(x - x^*) - \bar{\theta}\, f^T(x) M(x - x^*) - \bar{\theta}(x - x^*)^T M\, f(x) \\ &\qquad + \bar{\theta}^2\, f^T(x)\, Mf(x). \end{aligned}$$

Hence, $T(x) = [1 - R(x)]^{\frac{1}{2}}$, where

$$R(x) := \frac{\bar{\theta}\, f^T(x) M(x - x^*) + \bar{\theta}(x - x^*)^T M\, f(x) - \bar{\theta}^2\, f^T(x) M\, f(x)}{(x - x^*)^T M(x - x^*)}. \tag{3.15}$$

In order to determine an expression for $\bar{\theta}$, we approximate f about x^* with a linear mapping as follows:

$$f(x) \cong f(x^*) + \nabla f(x^*)(x - x^*) = M(x - x^*).$$

Let V_x denote the error in linearly approximating $f(x)$ about x^*, i.e.

$$V_x = f(x) - M(x - x^*).$$

Substituting $M(\bar{x} - x^*) + V_{\bar{x}} = M(x - \bar{\theta}\, f(x) - x^*) + V_{\bar{x}}$ for $f(\bar{x})$ in (3.14) implies that $f^T(x)M(x - x^*) - \bar{\theta}\, f^T(x)Mf(x) + f^T(x)V_{\bar{x}} = 0$, and, hence, that

$$\bar{\theta} = \frac{f^T(x)\, M(x - x^*) + f^T(x) V_{\bar{x}}}{f^T(x) Mf(x)}.$$

Substituting for $\bar{\theta}$ in (3.15) and simplifying, we have that

$$R(x) = \{[f^T(x)M(x - x^*)][(x - x^*)^T Mf(x)] + [(x - x^*)^T Mf(x)][f^T(x)V_{\bar{x}}]$$

$$-[f^T(x)M(x - x^*)][f^T(x)V_{\bar{x}}] - [f^T(x)V_{\bar{x}}]^2\} \div \qquad (3.16)$$

$$\{[f^T(x)Mf(x)][(x - x^*)^T M(x - x^*)]\}.$$

Since $r = \sup_{x \varepsilon S_\varepsilon} T(x) = \sup_{x \varepsilon S_\varepsilon} [1 - R(x)]^{1/2} = [1 - \inf_{x \varepsilon S_\varepsilon} R(x)]^{1/2}$,

$r<1$ if and only if $\inf_{x \varepsilon S_\varepsilon} R(x) > 0$. We will, therefore, show that $\inf_{x \varepsilon S_\varepsilon} R(x) > 0$.

Substituting $M(x - x^*) + V_x$ for $f(x)$ in the first term in the numerator of (3.16) and collecting terms gives

$$R(x) = \frac{[(x - x^*)^T M^T M(x - x^*)][(x - x^*)^T M^2(x - x^*)] - E(x)}{[f^T(x)\ Mf(x)][(x - x^*)^T M(x - x^*)]},$$

where the error term $E(x)$ contains all terms involving V_x and $V_{\bar{x}}$, and is given by

$$-E(x) = [V_x^T\ M(x - x^*)][(x - x^*)^T\ M^2(x - x^*)]$$

$$+ [(x - x^*)^T M^T M(x - x^*)][(x - x^*)^T MV_x] + [V_x^T\ M(x - x^*)][(x - x^*)^T MV_x]$$

$$+ [(x - x^*)^T Mf(x)][f^T(x)V_{\bar{x}}] - [f^T(x)M(x - x^*)][f^T(x)V_{\bar{x}}]$$

$$- [f^T(x)V_{\bar{x}}]^2.$$

$E(x)$ can be bounded from above using the triangle inequality, Cauchy's inequality, and the fact that the matrix norm $||A||$ satisfies $||Ax|| \leq ||A||\,||x||$ for every vector x:

$$E(x) \leq |E(x)| \leq \| V_x \| \, \| M \| \, \| M^2 \| \, \| x - x^* \|^3 + \| V_x \| \, \| M \|^2 \, \| M^T \| \, \| x - x^* \|^3$$

$$+ \| V_x \|^2 \, \| M \|^2 \, \| x - x^* \|^2 + \| V_{\bar{x}} \| \, \| M \| \, \| f(x) \|^2 \| x - x^* \|$$

$$+ \| V_{\bar{x}} \| \, \| M \| \, \| f(x) \|^2 \| x - x^* \| + \| V_{\bar{x}} \|^2 \, \| f(x) \|^2 .$$

By Lemma 3.6, we have that for any $x \varepsilon S_\varepsilon$,

$$\| V_x \| \leq c_1 \| x - x^* \|^2 ,$$

$$\| f(x) \| \leq c_2 \| x - x^* \|, \text{ and}$$

$$\| V_{\bar{x}} \| \leq c_4 \| x - x^* \|^2 ,$$

where $c_1 \geq 0$, $c_2 > 0$ and $c_4 > 0$.

Thus, $E(x) \leq K_1 \| x - x^* \|^5 + K_2 \| x - x^* \|^5 + K_3 \| x - x^* \|^6 + K_4 \| x - x^* \|^5 +$

$$+ K_5 \| x - x^* \|^5 + K_6 \| x - x^* \|^6$$

$$\leq K \| x - x^* \|^5 ,$$

where the $K_i \geq 0$, and, hence, $K := K_1 + K_2 + K_3 + K_4 + K_5 + K_6 \geq 0$.

Thus, $R(x) \geq \dfrac{[(x - x^*)^T M^T M (x - x^*)][(x - x^*)^T M^2 (x - x^*)] - K \| x - x^* \|^5}{[f^T(x) M f(x)][(x - x^*)^T M (x - x^*)]}$.

Dividing the numerator and denominator by $\| x - x^* \|^2 [(x - x^*)^T M^T M (x - x^*)]$ gives

$$\inf_{x\varepsilon S_\varepsilon} R(x) \geq \inf_{x\varepsilon S_\varepsilon}\left\{\frac{\dfrac{(x-x^*)^T M^2(x-x^*)}{(x-x^*)^T M(x-x^*)} - \dfrac{K\|x-x^*\|^3}{(x-x^*)^T M^T M(x-x^*)}}{\dfrac{f^T(x)Mf(x)}{(x-x^*)^T M^T M(x-x^*)}}\right\}$$

$$\geq \frac{\inf\limits_{x\varepsilon X_\varepsilon}\left\{\dfrac{(x-x^*)^T M^2(x-x^*)}{(x-x^*)^T(x-x^*)}\right\} - \sup\limits_{x\varepsilon S_\varepsilon}\left\{\dfrac{K\|x-x^*\|^3}{(x-x^*)^T M^T M(x-x^*)}\right\}}{\sup\limits_{x\varepsilon S_\varepsilon}\left\{\dfrac{f^T(x)Mf(x)}{(x-x^*)^T M^T M(x-x^*)}\right\}}$$

Considering each of these three expressions separately, we have:

$$\inf_{x\varepsilon S_\varepsilon}\left\{\frac{(x-x^*)^T M^2(x-x^*)}{(x-x^*)^T M(x-x^*)}\right\} \geq \frac{\inf\limits_{x\varepsilon S_\varepsilon}\left\{\dfrac{(x-x^*)^T M^2(x-x^*)}{(x-x^*)(x-x^*)}\right\}}{\sup\limits_{x\varepsilon S_\varepsilon}\left\{\dfrac{(x-x^*)^T M(x-x^*)}{(x-x^*)^T(x-x^*)}\right\}} = \frac{\lambda_{min}(\widehat{M^2})}{\lambda_{max}(\hat{M})} > 0$$

since M^2 and $\hat{M}$ are positive definite;

$$\sup_{x\varepsilon S_\varepsilon}\left\{\frac{K\|x-x^*\|^3}{(x-x^*)^T M^T M(x-x^*)}\right\} = \sup_{x\varepsilon S_\varepsilon}\left\{\frac{K\|x-x^*\|}{\dfrac{(x-x^*)^T M^T M(x-x^*)}{(x-x^*)^T(x-x^*)}}\right\}$$

$$\leq \frac{\sup\limits_{x\varepsilon S_\varepsilon} K\|x-x^*\|}{\inf\limits_{x\varepsilon S_\varepsilon}\left\{\dfrac{(x-x^*)^T M^T M(x-x^*)}{(x-x^*)^T(x-x^*)}\right\}}$$

$$= \frac{K\varepsilon}{\lambda_{min}(M^T M)} = b\varepsilon,$$

where $b := K/\lambda_{min}(M^T M) \geq 0$; and

$$\sup_{x \varepsilon S_\varepsilon}\left\{\frac{f^T(x)Mf(x)}{(x - x^*)^T M^T M(x - x^*)}\right\} = \sup_{x \varepsilon S_\varepsilon}\left\{\frac{[M(x - x^*) + v_x]^T M[M(x - x^*) + v_x]}{(x - x^*)^T M^T M(x - x^*)}\right\}$$

$$= \sup_{x \varepsilon S_\varepsilon}\left\{\frac{[M(x - x^*)]^T M[M(x - x^*)] + v_x^T M^2(x - x^*) + (x - x^*)^T M^T M v_x + v_x^T M v_x}{(x - x^*)^T M^T M(x - x^*)}\right\}$$

$$\leq \sup_{x \varepsilon S_\varepsilon}\left\{\frac{[(M(x - x^*)]^T M[M(x - x^*)]}{[M(x - x^*)]^T[M(x - x^*)]} + \sup_{x \varepsilon S_\varepsilon} \frac{v_x^T M^2(x - x^*) + (x - x^*)^T M^T M v_x + v_x^T M v_x}{(x - x^*)^T M^T M(x - x^*)}\right\}$$

$$\leq \lambda_{max}(\hat{M}) + \sup_{x \varepsilon S_\varepsilon}\left\{\frac{\| v_x\| \, \| M^2\| \, \| x - x^*\| + \| x - x^*\| \, \| M^T M\| \, \| v_x\| + \| v_x\| \, \|M\| \, \|v_x\|}{(x - x^*)^T M^T M(x - x^*)}\right\}$$

$$\leq \lambda_{max}(\hat{M}) = \sup_{x \varepsilon S_\varepsilon}\left\{\frac{c\| x - x^*\|^3}{(x - x^*)^T M^T M(x - x^*)}\right\}$$

$$\leq \lambda_{max}(\hat{M}) + \frac{\sup_{x \varepsilon S_\varepsilon} c\| x - x^*\|}{\inf_{x \varepsilon S_\varepsilon}\left\{\frac{(x - x^*)^T M^T M(x - x^*)}{(x - x^*)^T(x - x^*)}\right\}}$$

$$= \lambda_{max}(\hat{M}) + \frac{c\varepsilon}{\lambda_{min}(M^T M)} = \lambda_{max}(\hat{M}) + a\varepsilon$$

where $c := c_1\| M^2\| + c_1\| M^T M\| + c_1^2\| M\| \geq 0$, and, hence, $a := c/\lambda_{min}(M^T M) \geq 0$.

Combining these inequalities gives

$$\inf_{x \varepsilon S_\varepsilon} R(x) \geq \frac{\left\{\frac{\lambda_{min}(\widehat{M^2})}{\lambda_{max}(\hat{M})}\right\} - b\varepsilon}{\lambda_{max}(\hat{M}) + a\varepsilon},$$

which is greater than zero if ε is sufficiently small, since the denominator is positive, $\lambda_{min}(\widehat{M^2})/\lambda_{max}(\hat{M}) > 0$, and $b \geq 0$. ∎

Lemma 3.4

If f is uniformly monotone, then, for a given $x \neq x^*$, there exists a unique $\bar{\theta} > 0$ satisfying $f^T(x)f(x - \bar{\theta}f(x)) = 0$.

Proof

$\bar{\theta} \geq 0$ solves the one-dimensional variational inequality problem

$$[x - \theta f(x) - (x - \bar{\theta}f(x))]^T f(x - \bar{\theta}f(x)) \geq 0 \qquad \text{for every } \theta \geq 0,$$

i.e., $\bar{\theta}$ satisfies

$$-(\theta - \bar{\theta})f^T(x)f(x - \bar{\theta}f(x)) \geq 0 \qquad \text{for every } \theta \geq 0.$$

Thus, $\bar{\theta}$ solves $VI(g,R^1)$, where $g(\theta) := - f^T(x)f(x-\theta f(x))$.
The existence and uniqueness of $\bar{\theta}$ follow, because, for a given x, g is uniformly monotone with modulus of monotonicity $\alpha\|f(x)\|^2$:

$$\begin{aligned}
(\theta_2-\theta_1)[g(\theta_2)-g(\theta_1)] &= (\theta_2-\theta_1)f^T(x)[f(x-\theta_1 f(x)) - f(x-\theta_2 f(x))] \\
&= [(x-\theta_1 f(x)) - (x-\theta_2 f(x))]^T[f(x-\theta_1 f(x))-f(x-\theta_2 f(x))] \\
&\geq \alpha\| x - \theta_1 f(x) - x + \theta_2 f(x)\|^2 \\
&= \alpha(\theta_2 - \theta_1)^2 \|f(x)\|^2 \\
&= [\alpha\|f(x)\|^2](\theta_2 - \theta_1)^2. \qquad (3.17)
\end{aligned}$$

Moreover, $\bar{\theta}$ is positive, because $x \neq x^*$ implies that $g(0) = -\|f(x)\|^2 \neq 0$. ∎

Lemma 3.5

If f is uniformly monotone with modulus of monotonicity α, then for any $x \neq x^*$, the unique $\bar{\theta} > 0$ for which $f^T(x)f(x - \bar{\theta}f(x)) = 0$ satisfies $\bar{\theta} \leq \frac{1}{\alpha}$.

Proof

By setting $\theta_2 = \bar{\theta}$ and $\theta_1 = 0$, (3.17) reduces to

$$\bar{\theta}f^T(x)[f(x) - f(x - \bar{\theta}f(x))] \geq \alpha\bar{\theta}^2 f^T(x)f(x),$$

or, since $\bar{\theta} > 0$ and $f^T(x)f(x - \bar{\theta}f(x)) = 0$,

$$f^T(x)f(x) \geq \alpha\bar{\theta}f^T(x)f(x).$$

Because $x \neq x^*$, $f(x) \neq 0$, so $\alpha\bar{\theta} \leq 1$. Therefore, since $\alpha > 0$, $\bar{\theta} \leq \frac{1}{\alpha}$. ∎

Lemma 3.6

Assume that f has a second Gateaux derivative on the open set $S: = \{x \mid \|x - x^*\| < 1\}$. Let $\bar{x} = x - \bar{\theta}f(x)$, where $\bar{\theta}$ is chosen so that $f^T(x)f(\bar{x}) = 0$. Let $V_x = f(x) - M(x - x^*)$, let $V_{\bar{x}} = f(\bar{x}) - M(\bar{x} - x^*)$, and let $\|\cdot\|$ denote the $\hat{M}$ norm.

Then, for $i = 1,2,3,4$, there exist constants $c_i \geq 0$ that satisfy the following conditions for any $x \in S$:

(i) $\quad \|V_x\| \leq c_1\|x - x^*\|^2$;

(ii) $\| f(x) \| \leq c_2 \|x - x^*\|$;

(iii) $\| \bar{x} - x^* \| \leq c_3 \| x - x^* \|$; and

(iv) $\| v_{\bar{x}} \| \leq c_4 \| x - x^* \|^2$.

Proof

(i) $\| v_x \| = \| f(x) - M(x - x^*) \|$

$$= \| f(x) - f(x^*) - \nabla f(x^*)(x - x^*) \|$$

$$\leq \sup_{0 \leq t \leq 1} \| \nabla^2 f[x^* + t(x - x^*)] \| \; \| x - x^* \|^2$$

$$\leq \left\{ \sup_{x \in S} \left\{ \sup_{0 \leq t \leq 1} \| \nabla^2 f[x^* + t(x - x^*)] \| \right\} \right\} \| x - x^* \|^2$$

$$= c_1 \| x - x^* \|^2,$$

where the first inequality follows from an extended mean value theorem stated as Theorem 3.3.6 in Ortega and Rheinboldt [1970]. Clearly, $c_1 \geq 0$.

(ii) $\| f(x) \| = \| M(x - x^*) + v_x \|$

$$\leq \| M \| \; \| x - x^* \| + \| v_x \|$$

$$\leq \| M \| \; \| x - x^* \| + c_1 \; \| x - x^* \|^2 \qquad \text{by (i)}$$

$$\leq (\| M \| + c_1) \| x - x^* \| \qquad \text{since } \| x - x^* \| < 1$$

$$= c_2 \| x - x^* \|.$$

$c_2 := \|M\| + c_1 > 0$ because $\|M\| > 0$ and $c_1 \geq 0$.

(iii) $$\| \bar{x} - x^* \| = \| x - \bar{\theta} f(x) - x^* \|$$

$$\leq \| x - x^* \| + \bar{\theta} \; \| f(x) \|$$

$$\leq \| x - x^* \| + \frac{1}{\alpha} c_2 \| x - x^* \| \qquad \text{by Lemma 3.4 and (ii)}$$

$$= c_3 \| x - x^* \|.$$

$c_3 := 1 + \frac{1}{\alpha} c_2 > 0$ because $c_2 > 0$ and $\alpha > 0$.

(iv) $$\| V_{\bar{x}} \| = \| f(\bar{x}) - M(x - x^*) \|$$

$$= \| f(\bar{x}) - f(x^*) - \nabla f(x^*)(\bar{x} - x^*) \|$$

$$\leq \sup_{0 \leq t \leq 1} \| \nabla^2 f[x^* + t(\bar{x} - x^*)] \| \; \| \bar{x} - x^* \|^2$$

$$\leq \left\{ \sup_{\{\bar{x} : \|\bar{x} - x^*\| \leq c_3\}} \left\{ \underset{\text{as in (i)}}{\sup_{0 \leq t \leq 1} \| \nabla^2 f[x^* + t(\bar{x} - x^*)] \|} \right\} \right\} \|\bar{x} - x^*\|^2$$

$$= \bar{c}_4 \; \| \bar{x} - x^* \|^2$$

$$\leq \bar{c}_4 \, c_3^2 \; \| x - x^* \|^2 \qquad \text{by (ii)}$$

$$= c_4.$$

$c_4 \geq 0$ because $c_4 \geq 0$ and $c_3^2 \geq 0$. ∎

3.4 Scaling the Mapping of an Unconstrained Variational Inequality Problem

In this section, we consider a procedure for scaling the mapping of an unconstrained variational inequality problem that is to be solved by the generalized steepest descent algorithm. We first consider the problem VI(f,C) defined by the affine mapping $f(x) = Mx - b$. We show below that, by scaling either the rows or the columns of M in an appropriate manner before applying the generalized steepest descent algorithm, we can weaken, perhaps considerably, the convergence conditions that Theorem 3.4 imposes on M.

When $f(x) = Mx - b$, the unconstrained problem VI(f,C) is equivalent to the problem of finding a solution to the linear equation $Mx = b$. If A is a nonsingular nxn matrix, then the linear systems $Mx = b$ and $(AM)x = Ab$ are equivalent. We can, therefore, find the solution to $VI(f,R^n)$ by solving the equivalent problem $VI(Af, R^n)$, where $Af(x) = AMx-Ab$. The generalized steepest descent method will solve $VI(Af, R^n)$ if both AM and $(AM)^2$ are positive definite matrices. In particular, if we assume that M has positive diagonal entries, and let $D = \text{diag}(M)$, then D^{-1} is nonsingular. Therefore, the generalized steepest descent method will solve $VI(D^{-1}f, R^n)$ if $(D^{-1}M)$ and $(D^{-1}M)^2$ are both positive definite matrices, which is true, by Theorem 3.4, if for every $i = 1,2,\ldots,n$,

$$\sum_{j\neq i} |(D^{-1}M)_{ij}| < ct \text{ and } \sum_{j\neq 1} |(D^{-1}M)_{ji}| < ct, \tag{3.18}$$

$$\text{where } c = \sqrt{2}-1 \text{ and } t = \frac{\min\{(D^{-1}M)_{ii}^2 : i = 1,\ldots,n\}}{\max\{(D^{-1}M)_{ii} : i = 1,\ldots,n\}}.$$

Since D^{-1} is diagonal, and $(D^{-1})_{ii} = (M_{ii})^{-1}$, then for each i and j,

$$(D^{-1}M)_{ij} = M_{ij}/M_{ii},$$

and, in particular, for each i,

$$(D^{-1}M)_{ii} = 1.$$

The condition (3.18) can, therefore, be simplified:

for every i = 1,2,...,n,

$$\sum_{j\neq i} |M_{ij}| < c\, M_{ii} \text{ and } \sum_{j\neq i} \frac{|M_{ji}|}{M_{jj}} < c, \text{ where } c = \sqrt{2} - 1.$$

The above argument establishes the following result.

Theorem 3.6

Let $M = (M_{ij})$ be an nxn matrix with positive diagonal entries, and let $D^{-1} = [\mathrm{diag}(M)]^{-1}$. If for every i = 1,2,...,n,

$$\sum_{j\neq i} |M_{ij}| < c\, M_{ii} \text{ and } \sum_{j\neq i} \frac{|M_{ji}|}{M_{jj}} < c, \tag{3.19}$$

where $c = \sqrt{2} - 1$, then $(D^{-1}M)$ and $(D^{-1}M)^2$ are positive definite matrices.

The conditions that Theorem 3.6 imposes on M may be considerably less restrictive than the analogous conditions that Theorem 3.4 imposes on M; namely,

for every i = 1,2,...,n,

$$\sum_{j\neq i} |M_{ij}| < ct \text{ and } \sum_{j\neq i} |M_{ji}| < ct, \tag{3.20}$$

$$\text{where } c = \sqrt{2} - 1 \text{ and } t = \frac{\min\{M_{ii}^2 : i = 1,\ldots,n\}}{\max\{M_{ii} : i = 1,\ldots, n\}} .$$

The conditions on the row sums of M in (3.20) are at least as restrictive as those in (3.19), because $t \le M_{ii}$ for every $i = 1,2,\ldots,n$. This is also true for the column sum conditions: because $t \le \min\{M_{ii}: i = 1,\ldots,n\}$, $\sum_{j\neq i} |M_{ji}| < c \min\{M_{ii}: i = 1,\ldots,n\}$, which implies that

$$\sum_{j\neq i} \frac{|M_{ji}|}{M_{jj}} \le \sum_{j\neq i} \frac{|M_{ij}|}{\text{Min}\{M_{ii}:i=1,\ldots,n\}} \le c.$$ The conditions specified in (3.20) are equivalent to those given in (3.19) if, and only if, all of the diagonal entries of M are identical.

Let us return to Example 3.2 and consider the effect of scaling the matrix M, where

$$M = \begin{bmatrix} N & 0 & 0 \\ a & 1 & 0 \\ b & 0 & 1 \end{bmatrix} \cdot \text{ Here, } D^{-1} = \begin{bmatrix} 1/N & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ and } D^{-1}M = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & 0 & 1 \end{bmatrix}.$$

The conditions (3.19) reduce to the inequality

$$|a| + |b| < \sqrt{2} - 1.$$

In contrast, the conditions (3.20) for the unscaled problem are

$$|a| + |b| < \begin{cases} (\sqrt{2} - 1)\ N^2 & \text{if } 0 \le N \le 1 \\ (\sqrt{2} - 1)\ \frac{1}{N} & \text{if } N \ge 1 \end{cases} .$$

Consider these conditions on M for the unscaled problem for various values of N:

N	upper bound on $\lvert a\rvert + \lvert b\rvert$
$\frac{1}{4}$	$\frac{1}{16}\ (\sqrt{2} - 1)$
$\frac{1}{2}$	$\frac{1}{4}\ (\sqrt{2} - 1)$
1	$1\ (\sqrt{2} - 1)$
2	$\frac{1}{2}\ (\sqrt{2} - 1)$
10	$\frac{1}{10}\ (\sqrt{2} - 1)$

The upper bound on $|a| + |b|$ is tighter than the upperbound $(\sqrt{2} - 1)$ imposed on $|a| + |b|$ for the scaled problem unless $N = 1$, in which case the bounds are the same. As the value of N moves away from 1, the conditions imposed on $|a|$ and $|b|$ for the unscaled problem becomes increasingly stringent. (As $N \to 0$ or $N \to \infty$, the conditions (3.19) drive $|a|$ and $|b|$ to zero).

In Example 3.3, we see a similar trend.

Here,

$$M = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & 0 & N \end{bmatrix}. \text{ Hence, } D^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{N} \end{bmatrix}, \text{ and } D^{-1}M = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ \frac{b}{N} & 0 & 1 \end{bmatrix}.$$

In this case, the conditions (3.19 reduce to

$$|a| + \frac{|b|}{N} < \sqrt{2} - 1,$$

while the conditions (3.20) on the unscaled problem are

$$|a| + |b| < \begin{cases} (\sqrt{2} - 1)N^2 & \text{if} \quad 0 \le N \le 1 \\ (\sqrt{2} - 1)\frac{1}{N} & \text{if} \quad N \ge 1 \ . \end{cases}$$

Consider the two sets of conditions for several values of N:

N	Scaled Problem	Unscaled Problem
$\frac{1}{4}$	$\|a\| + 4\|b\| < \sqrt{2} - 1$	$\|a\| + \|b\| < \frac{1}{16}(\sqrt{2} - 1)$
$\frac{1}{2}$	$\|a\| + 2\|b\| < \sqrt{2} - 1$	$\|a\| + \|b\| < \frac{1}{4}(\sqrt{2} - 1)$
1	$\|a\| + \|b\| < \sqrt{2} - 1$	$\|a\| + \|b\| < \sqrt{2} - 1$
2	$\|a\| + \frac{1}{2}\|b\| < \sqrt{2} - 1$	$\|a\| + \|b\| < \frac{1}{2}(\sqrt{2} - 1)$
10	$\|a\| + \frac{1}{10}\|b\| < \sqrt{2} - 1$	$\|a\| + \|b\| < \frac{1}{10}(\sqrt{2} - 1)$

Again, the conditions on the unscaled problem become increasingly more stringent than those on the unconstrained problem as N moves away from 1.

Analogous results can be obtained by column-scaling the matrix M. An argument similar to the above shows that

if for every $i = 1,2,\ldots,n$,

$$\sum_{j \neq i} \frac{|M_{ij}|}{M_{jj}} < c \text{ and } \sum_{j \neq i} |M_{ji}| < c\, M_{ii}, \qquad \text{where } c = \sqrt{2} - 1. \tag{3.21}$$

then MD^{-1} and $(MD^{-1})^2$ are positive definite.

For a given problem, either the rows or the columns of M could be scaled in order to satisfy one of the sets of conditions that ensure convergence of the generalized steepest descent method. For a given matrix, one of these scaling procedures may define a matrix for which the algorithm will work, even if the other does not. If we column-scale the matrix of Example 3.2, then

$$MD^{-1} = \begin{bmatrix} 1 & 0 & 1 \\ a/N & 1 & 0 \\ b/N & 0 & 1 \end{bmatrix},$$

and conditions (3.21) reduce to

$$|a| + |b| < N.$$

Thus, in order to obtain the least restrictive conditions on M, it is better to row scale if $N < 1$, and column scale if $N > 1$.

By using the above row- or column-scaling procedure, it may be possible to transform a variational inequality problem defined by a nonmonotone affine map into a problem defined by a strictly monotone affine map. That is, $D^{-1}M$ or MD^{-1} may be positive definite, even if M is not. The following example illustrates such a situation.

<u>Example 3.4</u>

Let $M = \begin{bmatrix} 1 & 0.5 \\ 8 & 10 \end{bmatrix}$. M is not positive definite, since

$$\det \hat{M} = \det \begin{bmatrix} 1 & 4.25 \\ 4.25 & 10 \end{bmatrix} = -8.0625 < 0 .$$

However, $\widehat{D^{-1}M} = \begin{bmatrix} 1 & 0.5 \\ 0.8 & 1 \end{bmatrix}$ is positive definite, since

$$\det \widehat{D^{-1}M} = \det \begin{bmatrix} 1 & 0.65 \\ 0.65 & 1 \end{bmatrix} = 0.5775 > 0.$$

Note that column scaling does not produce a positive definite matrix, since $MD^{-1} = \begin{bmatrix} 1 & 0.05 \\ 8 & 1 \end{bmatrix}$, and

$$\det \widehat{MD^{-1}} = \det \begin{bmatrix} 1 & 4.025 \\ 4.025 & 1 \end{bmatrix} = -15.2 < 0.$$

Note also that $(D^{-1}M)^2 = \begin{bmatrix} 1.4 & 1 \\ 1.6 & 1.4 \end{bmatrix}$ is positive definite, since

$$\det \widehat{(D^{-1}M)^2} = \det \begin{bmatrix} 1.4 & 1.3 \\ 1.3 & 1.4 \end{bmatrix} = 0.27 > 0.$$

Consequently, any unconstrained problem defined by $f(x) = Mx-b$, where M is the above defined matrix, can be transformed, by row-scaling, into an equivalent problem that can be solved by the generalized steepest descent method, even though neither M nor M^2 is positive definite. ∎

The above scaling procedures can also be used to transform a nonlinear mapping into one that satisfies the convergence conditions given in Theorem 3.5 for the generalized steepest descent algorithm. The algorithm will converge in a neighborhood of the solution x^* if

(i) $D^{-1}f$ (or fD^{-1}) is uniformly monotone and twice Gateaux differentiable; and

(ii) $[D^{-1}\nabla f(x^*)]^2$ (or $[\ f(x^*)D^{-1}]^2$) is positive definite.

3.5 Generalized Descent Algorithms

The steepest algorithm for the unconstrained minimization problem,

$$\min_{x \in R^n} F(x),$$

generates a sequence of iterates $\{x^k\}$ by determining a point $x^{k+1} \in R^n$ that minimizes F in the direction $-\nabla F(x^k)$ from the previous iterate x^k. In contrast, general descent methods (or gradient methods) generate a sequence of iterates $\{x^k\}$ by determining a point $x^{k+1} \in R^n$ that minimizes F in the direction d_k from x^k, where d_k is any descent direction from $x^k \neq x^*$, i.e., $d_k^T \nabla F(x^k) < 0$. The set of descent directions for F from the point $x^k \neq x^*$ is given by

$$D(x^k) := \{-A_k \nabla F(x^k) : A_k \text{ is an nxn positive definite matrix}\}.$$

This general descent method reduces to the steepest descent method when $A_k = I$ for $k = 0,1,2,\ldots$. If $A_k = [\nabla^2 F(x^k)]^{-1}$ for $k = 0,1,2,\ldots$, then this method becomes a "damped" or "limited-step" Newton method. If F is uniformly convex and twice-continuously differentiable, then this modification of Newton's method (i.e., Newton's method with a minimizing step length) will produce iterates converging to the unique critical point of F. (See, for example, Ortega and Rheinboldt [1970].)

In this section, we analyze the convergence of gradient algorithms adapted to solve unconstrained variational inequality problems.

Let A_k for $k = 0,1,2,\ldots$ be a sequence of positive definite symmetric matrices. The generalized descent algorithm for the unconstrained inequality problem $VI(f,R^n)$ can be stated as follows:

Generalized Descent Algorithm

Step 0: Select $x^0 \varepsilon R^n$. Set $k = 0$.

Step 1: Compute $- A_k f(x^k)$.

If $A_k f(x^k) = 0$, stop: $x^k = x^*$.

Otherwise, go to Step 2.

Step 2: Find $x^{k+1} \varepsilon [x^k; - A_k f(x^k)]$ satisfying

$(x - x^{k+1})^T f(x^{k+1}) \geq 0$ for every $x \varepsilon [x^k; - A_k f(x^k)]$.

Go to Step 1 with $k = k + 1$. ∎

The following result summarizes the convergence properties for this algorithm when f is a strictly monotone affine mapping.

Theorem 3.7

Let M be a positive definite matrix, and $f(x) = Mx - b$. Let A_k for $k = 0,1,2,\ldots$ be a sequence of positive definite symmetric matrices, and let $\{x^k\}$ be the sequence of iterates generated by the generalized descent algorithm applied to $VI(f,C)$. Then,

(1) the steplength θ_k determined on the k^{th} iteration of the algorithm is

$$\theta_k = \frac{(Mx^k - b)^T A_k (Mx^k - b)}{(Mx^k - b)^T A_k \, M \, A_k (Mx^k - b)}; \tag{3.22}$$

(2) the iterates generated by the algorithm are guaranteed to contract to the solution x^* in $\hat{M}$ norm by a fixed contraction constant $r < 1$ if and only if

(i) $\inf_{k=0,1,\ldots} [\lambda_{min}(\widehat{MA_kM})] > 0$, and

(ii) $\inf_{k=0,1,\ldots} [\lambda_{min}(A_k)] > 0$; and

(3) the contraction constant r is bounded from above by

$$r' = \left\{ 1 - \frac{\inf_{k=0,1,\ldots} \lambda_{min}[(\hat{M}^{\frac{1}{2}})^{-1} (\widehat{MA_kM}) (\hat{M}^{\frac{1}{2}})^{-1}}{\sup_{k=0,1,\ldots} \lambda_{max}[A_k^{\frac{1}{2}} \hat{M} (A_k^{\frac{1}{2}})^T]} \right\}^{1/2} .$$

Proof

(1) Let x^k and x^{k+1} denote, respectively, the k^{th} and $(k+1)^{st}$ iterate, and let A_k denote the k^{th} scaling matrix.

Then,

$$x^{k+1} = x^k - \theta_k A_k f(x^k),$$

where θ_k is chosen so that x^{k+1} solves the one-dimensional variational inequality subproblem on $[x^k; - A_k f(x^k)]$.

As in the generalized steepest descent method, we can assume that $x^{k+1} \neq x^k$, because, if $x^{k+1} = x^k$, then $f(x^k) = 0$, and the algorithm would have terminated in Step 1 of the k^{th} iteration. Since x^{k+1} solves the unconstrained one-dimensional subproblem,

$$f^T(x^k) A_k f(x^{k+1}) = 0.$$

Substituting for $f(x^k)$, x^{k+1} and $f(x^{k+1})$ in this equation gives

$$(Mx^k - b)^T A_k [M(x^k - \theta_k A_k (Mx^k - b)) - b] = 0.$$

The last equality shows that θ_k is given by expression (3.22).

(2) The iterates generated by the algorithm contract in $\hat{M}$ normal to the solution $x^* = M^{-1} b$ if and only if there exists a real number $r \varepsilon [0,1)$ that is independent of x^k and satisfies $\| x^{k+1} - x^* \|_{\hat{M}} \leq r \varepsilon \| x^k - x^* \|_{\hat{M}}$ whenever $x^k \neq x^*$. Thus, we define

$$r := \sup_{\substack{k=0,1,2,\ldots \\ x^k \neq x^*}} T_{A_k}(x^k),$$

where

$$T_{A_k}(x^k) := \frac{\|x^{k+1} - x^*\|_{\hat{M}}}{\|x^k - x^*\|_{\hat{M}}} \quad \text{for } x^k \neq x^*.$$

As in the proof of Theorem 3.3, we obtain a simplified expression for $T_{A_k}(x^k)$:

$$T_{A_k}(x^k) = [1 - R_{A_k}(y^k)]^{\frac{1}{2}},$$

where $y^k = x^k - x^*$, and

$$R_{A_k}(y^k) := \frac{[(My^k)^T A_k (My^k)][(y^k)^T M A_k My^k]}{[(My^k)^T A_k M A_k (My^k)][(y^k)^T My^k]}.$$

Therefore, $r = \sup\limits_{\substack{x^k \neq x^* \\ k=0,1,\ldots}} T_{A_k}(x^k) = \sup\limits_{\substack{y^k \neq 0 \\ k=0,1,\ldots}} [1 - R_{A_k}(y^k)]^{\frac{1}{2}} = [1 - \inf\limits_{\substack{y^k \neq 0 \\ k=0,1,\ldots}} R_{A_k}(y^k)]^{\frac{1}{2}}.$

Consequently, r < 1 if and only if $\inf\limits_{\substack{y^k \neq 0 \\ k=0,1,..}} R_{A_k}(y^k) > 0$.

To complete the proof of (2), we show that we can guarantee that

$\inf\limits_{\substack{y^k \neq 0 \\ k=0,1,..}} R_{A_k}(y^k) > 0$ if and only if $\inf\limits_{k=0,1,..} \lambda_{min}(A_k) > 0$ and

$\inf\limits_{k=0,1,..} \lambda_{min}(\widehat{MA_k M}) > 0$.

If $\inf\limits_{k=0,1,...} \lambda_{min}(\widehat{MA_k M}) \leq 0$ or $\inf\limits_{k=0,1,.} \lambda_{min}(A_k) \leq 0$,

then there exists a sequence of nonzero vectors $\{y^k\}$ such that

$\inf\limits_{k=0,1..} (y^k)^T M A_k M y^k \leq 0$ or $\inf\limits_{k=01,..} (y^k)^T A_k y^k \leq 0$. In either case,

$\inf\limits_{\substack{y^k \neq 0 \\ k=0,1,...}} R_{A_k}(y^k) \leq 0$; i.e., $r \geq 1$.

If $\inf\limits_{k=0,1,..} \lambda_{min}(A_k) > 0$ and $\inf\limits_{k=0,1,..} \lambda_{min}(\widehat{MA_k M}) > 0$, then

$$\inf_{\substack{y^k \neq 0 \\ k=0,1,..}} R_{A_k}(y^k) \geq \frac{\inf\limits_{\substack{y^k \neq 0 \\ k=0,1,..}} \left\{ \dfrac{[(My^k)^T A_k (My^k)]}{(My^k)^T (My^k)} \right\} \inf\limits_{\substack{y^k \neq 0 \\ k=0,1,..}} \left\{ \dfrac{[(y^k)^T MA_k My^k]}{(y^k)^T y^k} \right\}}{\sup\limits_{\substack{y^k \neq 0 \\ k=0,1,..}} \left\{ \dfrac{[(My^k)^T A_k M A_k (My^k)]}{(My^k)^T (My^k)} \right\} \sup\limits_{\substack{y^k \neq 0 \\ k=0,1...}} \left\{ \dfrac{[(y^k)^T M y^k]}{(y^k)^T y^k} \right\}}$$

$$= \frac{\inf\limits_{k=0,1..} \lambda_{min}(A_k) \cdot \inf\limits_{k=0,1,..} \lambda_{min}(\widehat{MA_k M})}{\sup\limits_{k=0,1..} \lambda_{max}(A_k \hat{M} A_k) \cdot \lambda_{max}(\hat{M})}$$

$$> 0$$

because $\sup\limits_{k=0,1\ldots} \lambda_{max}(A_k \hat{M} A_k) > 0$ and $\lambda_{max}(\hat{M})>0$ by the positive definiteness of M.

(3) By an argument analogous to the argument in the proof of the Corollary to Theorem 3.3,

$$\inf_{\substack{y^k \neq 0 \\ k=0,1,..}} R_{A_k}(y^k) \geq \frac{\inf\limits_{\substack{y^k \neq 0 \\ k=0,1,..}} \left\{ \dfrac{(y^k)^T MA_k My^k}{(y^k)^T My^k} \right\}}{\sup\limits_{\substack{y^k \neq 0 \\ k=0,1,..}} \left\{ \dfrac{(My^k)^T A_k \hat{M} A_k (My^k)}{(My^k)^T A_k (My^k)} \right\}}$$

$$= \frac{\inf\limits_{k=0,1..} \lambda_{min}[(\hat{M}^{\frac{1}{2}})^{-T} (MA_k M)(\hat{M}^{\frac{1}{2}})^{-1}]}{\sup\limits_{k=0,1..} \lambda_{max}[A_k^{\frac{1}{2}} \hat{M} (A_k^{\frac{1}{2}})^T]} .$$

The result follows from this inequality and the fact that

$$r = [1 - \inf_{\substack{y^k \neq 0 \\ k=0,1\ldots}} R_{A_k}(y^k)]^{\frac{1}{2}}.$$ ∎

The iterates produced by the generalized descent algorithm will converge to the solution if we replace conditions (i) and (ii) of Theorem 3.7 with the conditions

(i') $\lim\limits_{k\to\infty} \inf[\lambda_{min}(\widehat{MA_k M})] > 0$, and

(ii') $\lim\limits_{k\to\infty} \inf[\lambda_{min}(A_k)] > 0$.

In this case, the iterates do not necessarily contract to the solution.

3.6 Concluding Remarks

In this chapter, we have analyzed the behavior of the steepest descent method for unconstrained convex minimization problems when it is generalized to solve unconstrained monotone variational inequality problems. The generalized steepest descent algorithm need not converge when applied to the problem $VI(f,R^n)$ if the Jacobian of f is not symmetric, even if f is a uniformly monotone mapping.

When $f(x) = Mx - b$ is a strictly monotone affine map, we show in Theorem 3.3 that the condition that M^2 is positive definite is necessary and sufficient for the iterates generated by the algorithm to contract to the solution of $VI(f,R^n)$. Theorem 3.4 establishes double diagonal dominance conditions on the matrix M that ensure that M and M^2 are positive definite, and, therefore, that the algorithm converges. In section 3.3, these results are extended to the problem $VI(f,R^n)$, where f is a uniformly monotone, but not necessarily affine, mapping.

In section 3.4, we describe a scaling procedure that allows a much wider class of affine maps to satisfy the convergence conditions of the algorithm. Finally, in section 3.5, we analyze the convergence properties of a class of generalized descent methods that allow movement in any "descent" direction from the previous iterate at each iteration.

CHAPTER 4

GENERALIZATIONS OF FIRST-ORDER APPROXIMATION METHODS

Many algorithms to solve nonlinear optimization problems and systems of nonlinear equations rely upon the fundamental idea of iteratively approximating the nonlinear function that defines the problem. In this chapter we analyze several variational inequality algorithms that generalize a class of these approximation methods.

Let $F:R^n \to R^1$ be a convex, continuously differentiable function, and let C be a closed convex subset of R^n. Consider the convex minimization problem

$$\min_{x \in C} F(x).$$

At each iteration, a first-order approximation algorithm approximates F by a function depending on the gradient of F. Linear approximation methods are classical examples of such methods. Given an iterate x^k, a linear approximation method generates the next iterate by using a linear approximation $F^k(x)$ to F about x^k given by

$$F^k(x) = F(x^k) + (x - x^k)^T \nabla F(x^k). \tag{4.1}$$

For example, the Frank-Wolfe method, which we formally state in Section 4.3, determines the solution v^k to the subproblem

$$\min_{x \in C} F^k(x).$$

The algorithm then chooses as its next iterate a point x^{k+1} that minimizes F on the line segment $[x^k, v^k]$.

A more accurate type of first-order approximation would replace the (constant) vector $\nabla F(x^k)$ by the (nonlinear) gradient $\nabla F(x)$, giving the approximation

$$F(x) \cong \hat{F}^k(x) := F(x^k) + (x - x^k)^T \nabla F(x). \qquad (4.2)$$

In this chapter, we will investigate variations of both of these first-order approximation schemes adapted to solve variational inequality problems. Sections 4.1 and 4.2 analyze variational inequality algorithms that generalize first-order approximation methods using the approximation (4.2). Our analysis of a "contracting ellipsoid" algorithm in Section 4.1 provides a geometrical framework within which to view a number of variational inequality algorithms. In section 4.2, we study a subgradient algorithm that solves a max-min problem that is equivalent to the variational inequality problem. This algorithm solves problems defined by monotone mappings; it does not require strict or uniform monotonicity. Section 4.3 discusses the use of a generalization of the Frank-Wolfe algorithm for solving variational inequality problems, and establishes the convergence of a modification of the generalized Frank-Wolfe method.

4.1 A Contracting Ellipsoid Algorithm and Its Interpretation

In this section we discuss a generalized first-order approximation algorithm for solving a variational inequality problem defined by the monotone mapping f. If f is the gradient mapping of a convex function $F:R^n \to R^1$; i.e., $[\nabla F(x)]^T = f(x)$ for every $x \in C$, then VI(f,C) is equivalent to the convex minimization problem

$$\min_{x \in C} F(x).$$

Consider an algorithm to minimize F over C that successively minimizes the approximation $\hat{F}^k(x) := F(x^k) + \nabla F(x)(x - x^k)$ to F given in (4.2). That is, the algorithm generates a sequence of iterates $\{x^k\}$ by the recursion

$$x^{k+1} = \underset{x \varepsilon C}{\operatorname{argmin}} \ \hat{F}^k(x), \quad k=0,1,\ldots$$

or, equivalently,

$$x^{k+1} = \underset{x \varepsilon C}{\operatorname{argmin}} \ \nabla F(x)(x - x^k), \quad k=0,1,\ldots .$$

By replacing $[\nabla F]^T$ with the mapping f, we obtain the following algorithm that is applicable to any variational inequality problem.

Contracting Ellipsoid Algorithm

Step 0: Select $x^0 \varepsilon C$. Set $k = 0$.

Step 1: Select $x^{k+1} \varepsilon \underset{x \varepsilon C}{\operatorname{argmin}} \ (x - x^k)^T f(x)$.

If $x^{k+1} = x^k$, then stop: $x^k = x^*$.

Otherwise, return to Step 1 with $k = k + 1$. ∎

(The name of the algorithm is motivated by the fact that for unconstrained problems with certain affine maps, the algorithm produces a sequence of ellipsoids that contract to the solution. This algorithm should not be confused with Khachiyan's [1979] ellipsoid algorithm for linear programming.)

In this section, we analyze this general algorithm. To motivate the analysis of the algorithm for the problem VI(f,C), we first consider, in section 4.1.1, the use of the algorithm for unconstrained variational inequality problems defined by affine maps. In this simplified problem setting, we describe the geometry of the algorithm and analyze its convergence properties. Section 4.1.2 extends these results to constrained problems defined by affine maps. Section 4.1.3 analyzes the algorithm for the constrained variational inequality problem defined by a nonlinear, strictly monotone mapping f. In section 4.1.4, we discuss the role of symmetry of the Jacobian of f in the convergence of the contracting ellipsoid method and the generalized steepest descent method, and compare the convergence conditions for these two algorithms. Finally, Section 4.1.5 discusses relationships between the contracting ellipsoid method and a number of well-known algorithms for variational inequality problems.

4.1.1 Unconstrained Problems with Affine Maps

In this subsection, we restrict our attention to the unconstrained variational inequality problem defined by a strictly monotone affine map. That is, we assume that $f(x) = Mx - b$, where M is a positive definite nxn matrix.

In this case, the minimization subproblem

$$\min_{x \in R^n} (x - x^k)^T f(x)$$

is a strictly convex quadratic programming problem. The strict convexity of the objective function ensures that the first order optimality conditions are both necessary and sufficient for x^{k+1} to be the unique solution to the subproblem. That is, given x^k, the next iterate is the unique $x^{k+1} \varepsilon R^n$ satisfying

$$[(x^{k+1} - x^k)^T M + (Mx^{k+1} - b)^T]\ (x - x^{k+1})$$

$$= (x - x^{k+1})^T\ [(M + M^T)x^{k+1} - (M^T x^k + b)] \geq 0 \quad \text{for every } x \in R^n.$$

Because this variational inequality subproblem is unconstrained, the solution x^{k+1} must be a zero of the mapping

$$(M + M^T)x - (M^T x^k + b)$$

that defines the subproblem. Hence, x^{k+1} is given by

$$x^{k+1} = S^{-1}(M^T x^k + b), \quad k = 0,1,2,\ldots\ , \tag{4.3}$$

where $S = M + M^T$; or, equivalently,

$$x^{k+1} = x^k - S^{-1}(Mx^k - b)$$

$$= x^k - S^{-1}f(x^k), \quad k = 0,1,\ldots\ .$$

Before proceeding with a convergence analysis, we illustrate the mechanics of the algorithm in an example.

Example 4.1

Let $M = \begin{bmatrix} 1 & 2 \\ -2 & 4 \end{bmatrix}$ and $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. The algorithm generates iterates by the relation

$$x^{k+1} = S^{-1}(M^T x^k + b), \qquad k = 0,1,\ldots,$$

where $S = M + M^T = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$. The solution x^* is $M^{-1}b = \begin{pmatrix} 1/4 \\ 3/8 \end{pmatrix}$.

Let $x^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then $x^1 = \begin{pmatrix} 1 \\ 3/8 \end{pmatrix}$, $x^2 = \begin{pmatrix} 5/8 \\ 9/16 \end{pmatrix}$, $x^3 = \begin{pmatrix} 1/4 \\ 9/16 \end{pmatrix}$,

$x^4 = \begin{pmatrix} 1/16 \\ 15/32 \end{pmatrix}$, $x^5 = \begin{pmatrix} 1/16 \\ 3/8 \end{pmatrix}$, $x^6 = \begin{pmatrix} 5/32 \\ 21/64 \end{pmatrix}$, ...

Figure 4.1 illustrates a sequence of iterates $\{x^k\}$ as well as a sequence of ellipses $\{E_o^k\}$ which we now describe.

Because f is affine, the behavior of the algorithm can be described geometrically in terms of the level sets of the objective function of the k^{th} subproblem, $\min_{x \varepsilon R^n} (x - x^k)^T(Mx - b)$. The level set given by

$$E_\alpha^k := \{x : (x - x^k)^T(Mx - b) \leq \alpha\}$$

is an ellipsoid centered about the point that minimizes the objective function of the k^{th} subproblem. Consequently, each of the ellipsoidal level sets is centered about the next iterate,

$$x^{k+1} = S^{-1}(M^T x^k + b).$$

The level set E_0^k is of particular interest. Note that ∂E_0^k (the boundary of E_0^k) contains both the k^{th} iterate x^k and the solution x^* to the problem (since $Mx^* - b = 0$). Hence, the point x^{k+1} is equidistant, with respect to the $\hat{M}$ norm, from x^k and from x^*. Because the ellipses E_0^k are defined by the same matrix $\hat{M}$, they have the same structure and orientation. E_0^k also has the same structure and orientation as the ellipse about the solution: $E = \{x : (x - x^*)^T\hat{M}(x - x^*) = c\}$. Note also that the chord joining x^k to any point x on ∂E_0^k

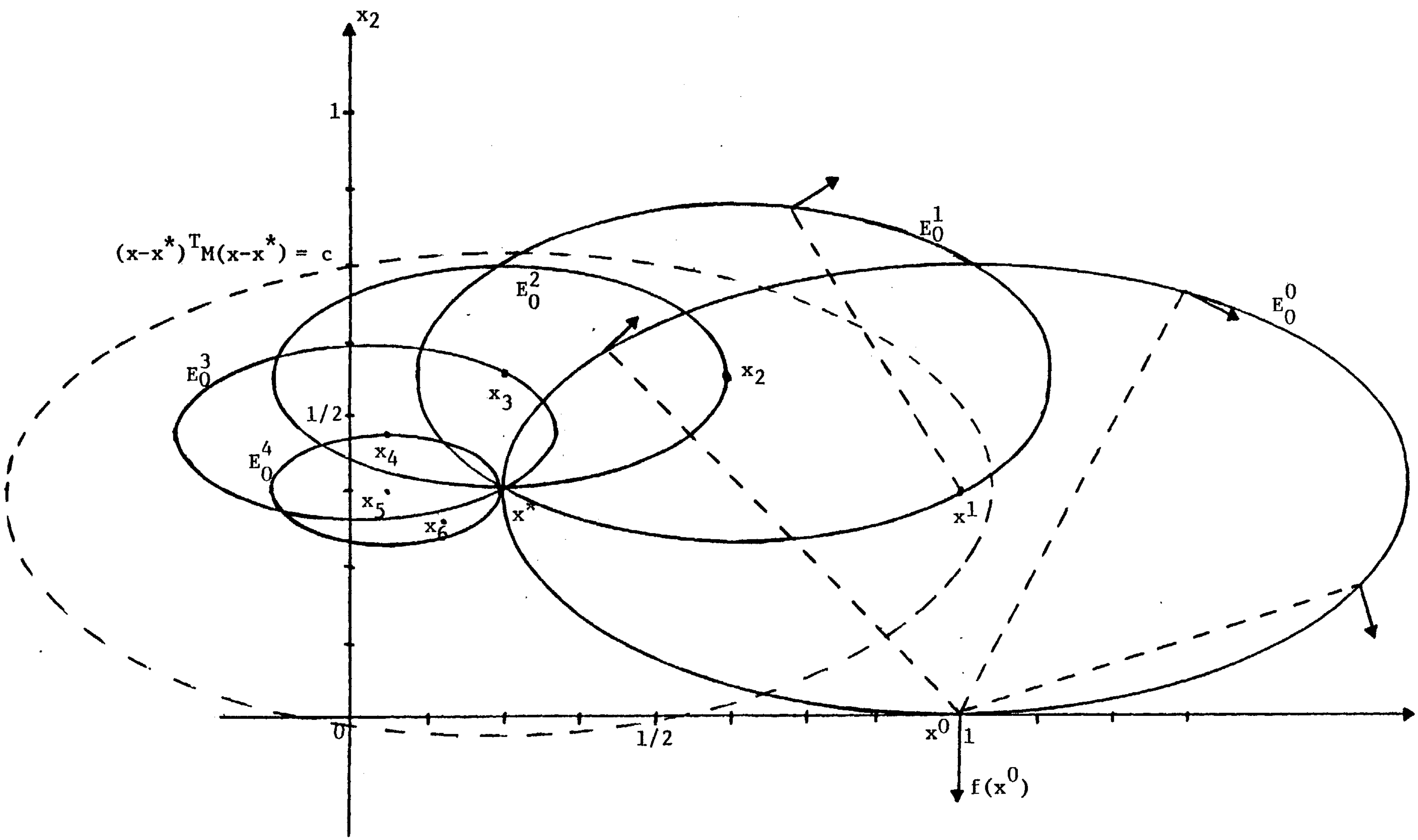


Figure 4.1 The Contracting Ellipsoid Method Solves An Unconstrained Affine Variational Inequality Problem by Generating a Sequence of Ellipsoids that Contract to the Solution x^*.

is orthogonal to the vector $f(x)$. This is true because, by definition of E_0^k, if $x \varepsilon \partial E_0^k$, then $(x - x^k)^T f(x) = 0$. This observation describes the relationship between the vector field defined by f and the ellipsoidal level sets.

The following theorem summarizes the convergence properties of the contracting ellipsoid algorithm for unconstrained problems defined by affine maps.

<u>Theorem 4.1</u>

If $f(x) = Mx - b$ and M is positive definite, then the sequence of iterates generated by the contracting ellipsoid algorithm converges to the solution $x^* = M^{-1}b$ if and only if the spectral radius $\rho(S^{-1}M^T)$ of the matrix $S^{-1}M^T$ is less than one.

Furthermore, for some norm $\|\cdot\|$, the algorithm converges linearly, with convergence ratio $\|S^{-1}M^T\|$.

<u>Proof</u>

From (4.3), $x^{k+1} = S^{-1}(M^T x^k + b)$, and, because the problem is unconstrained, $x^* = M^{-1}b$. Thus,

$$\begin{aligned} x^{k+1} - x^* &= S^{-1}[M^T x^k + b - (M + M^T)M^{-1}b] \\ &= S^{-1}M^T(x^k - x^*). \\ &= (S^{-1}M^T)^{k+1}(x^0 - x^*). \end{aligned}$$

The matrix $(S^{-1}M^T)^k$ approaches 0 as $k \to \infty$ if and only if $\rho(S^{-1}M^T) < 1$. (See, for example, Ortega and Rheinboldt [1970].) Hence, the sequence $\{x^k\}$ converges to x^* if and only if $\rho(S^{-1}M^T) < 1$.

Since $\rho(S^{-1}M^T) < 1$, there exists a norm $\|\cdot\|$ satisfying $\|S^{-1}M^T\| < 1$. By Cauchy's inequality, $\|x^{k+1} - x^*\| = \|S^{-1}M^T(x^k - x^*)\| \leq \|S^{-1}M^T\| \; \|x^k - x^*\|$ for each $k = 0,1,\ldots$, and, hence, the algorithm converges linearly, with convergence ratio $\|S^{-1}M^T\|$. ∎

The following lemma states several conditions that are equivalent to the condition $\rho(S^{-1}M^T) < 1$. In addition, the lemma shows that $\rho(S^{-1}M^T) < 1$, and, hence, the algorithm converges, whenever M^2 is positive definite. Consequently, if M satisfies the diagonal dominance conditions stated in Theorem 3.4, then the algorithm will converge. The row and column scaling procedures discussed in Section 3.4 can also be used in this setting to transform the matrix M into a matrix satisfying the conditions of Theorem 3.4.

Lemma 4.1

Let M be a positive definite matrix.

Then, (1) the following conditions are equivalent:

(i) $\rho(S^{-1}M^T) < 1$;

(ii) $\rho(MS^{-1}) < 1$;

(iii) $\rho[(M^{-T}M + I)^{-1}] < 1$;

(iv) $\rho[(M^TM^{-1} + I)^{-1}] < 1$;

(v) $\min\limits_{\lambda\varepsilon\lambda(M^{-T}M)} |\lambda + 1| > 1$; and

(vi) $\min_{\lambda \varepsilon \lambda(M^T M^{-1})} |\lambda + 1| > 1$; and

(2) if M^2 is positive definite, then $\rho(S^{-1}M^T) < 1$.

Proof

(1) First note that $(S^{-1}M^T)^T = MS^{-1}$, and $(M^{-T}M)^T = M^T M^{-1}$. The following equivalences are, therefore, a consequence of the fact that a matrix and its transpose have the same eigenvalues: (i) $\leftrightarrow$ (ii); (iii) $\leftrightarrow$ (iv); and (v) $\leftrightarrow$ (vi).

Conditions (i) and (ii) are equivalent because $(S^{-1}M^T) = (M^{-T}S)^{-1} = (M^{-T}M + I)^{-1}$.

Conditions (iii) and (v) are equialent because

$$\rho[(M^{-T}M + I)^{-1}] = \frac{1}{\min\{|\lambda| : \lambda \varepsilon \lambda(M^{-T}M+I)\}}$$

$$= \frac{1}{\min\{|\lambda + 1| : \lambda \varepsilon \lambda(M^{-T}M)\}}$$

$$< 1$$

if and only if $\min\{|\lambda + 1| : \lambda \varepsilon \lambda(M^{-T}M)\} > 1$.

(2) The matrix M^2 is positive definite if and only if $M^{-T}M$ is positive definite, since $M^2 = M^T(M^{-T}M)M$, and M is nonsingular. If $M^{-T}M$ is positive definite, then for every $\lambda \varepsilon \lambda(M^{-T}M)$, $\text{Re}\lambda > 0$, and, hence,

$$|\lambda + 1| = [(1 + \mathrm{Re}\lambda)^2 + (\mathrm{Im}\lambda)^2]^{1/2} \geq |1 + \mathrm{Re}\lambda| > 1.$$

By (1), (v) holds, and, hence (i) holds. ∎

Let us return to Example 4.1. The iterates produced by the algorithm are guaranteed to converge to the solution $x^* = \begin{pmatrix} 1/4 \\ 3/8 \end{pmatrix}$ because $\rho(S^{-1}M^T) = \rho \begin{bmatrix} 1/2 & -1 \\ 1/4 & 1/2 \end{bmatrix} = \frac{\sqrt{2}}{2} < 1$. This example illustrates that the condition that M^2 be positive definite is not a necessary condition for convergence: for this problem,

$$M^2 = \begin{bmatrix} -3 & 10 \\ -10 & 12 \end{bmatrix} \text{ is not positive definite.}$$

The geometrical interpretation of the algorithm discussed in Example 4.1 extends to all unconstrained problems defined by affine maps. The contracting ellipsoid method generates a sequence of ellipsoidal level sets for any such problem. Recall from our previous discussion that for each k, the ellipsoid E_0^k is centered about the point x^{k+1}, and that $x^* \varepsilon \partial E_0^k$. In addition, the ellipsoids $\{E_0^k\}$ all have the same structure. Therefore, if M satisfies $\rho(S^{-1}M^T) < 1$, then the fact that the sequence of ellipsoid centers x^{k+1} contracts in some norm to x^*, and that x^* is on the boundary of each E_0^k, ensures that the sequence of ellipsoids converges to the point x^*.

The distance with respect to the S norm from the center x^{k+1} of the k^{th} ellipsoid E_0^k to any point on its boundary is equal to $\| x^{k+1} - x^* \|_S$. Therefore, if M satisfies $\| S^{-1}M^T \|_S < 1$, and, hence, the sequence of iterates x^k contracts to the solution x^* in S norm, then the ellipsoids must contract to the solution in S norm.

The above observations establish the following result.

Theorem 4.2

Let $f(x) = Mx - b$, and $S = M + M^T$. If M is positive definite, and $\rho(S^{-1}M^T) < 1$, then the sequence of ellipsoids $\{E_0^k\}$ generated by the algorithm converges to the solution $x^* = M^{-1}b$. If, moreoever, $\| S^{-1}M^T \|_S < 1$, then the sequence $\{E_0^k\}$ contracts to the solution x^*.

4.2.2 Constrained Problems with Affine Maps

In this section, we extend the analysis of the previous section to the constrained problem VI(f,C), where f is a strictly monotone affine mapping and C is a closed, convex, nonempty subset of R^n. We let $f(x) = Mx - b$, where M is a positive definite nxn matrix.

Because f is affine, the minimization subproblem

$$\min_{x \in C} \quad (x - x^k)^T f(x) \tag{4.5}$$

is a strictly convex quadratic programming problem. Thus, the contracting ellipsoid algorithm solves the problem VI(f,C) by solving a sequence of quadratic programs. The work involved in this algorithm is, therefore, comparable to that of a projection algorithm, which also requires the solution of a sequence of quadratic programming problems.

The necessary and sufficient conditions for x^{k+1} to solve the k^{th} quadratic programming subproblem are

$$(x - x^{k+1})^T[(M + M^T)x^{k+1} - (M^Tx^k + b)] \geq 0 \qquad \text{for every } x\varepsilon C.$$

Hence, the subproblem is a variational inequality problem defined over C by the affine map

$$g(x,x^k) := (M + M^T)x - (M^Tx^k + b).$$

An alternative interpretation of the algorithm is, therefore, that it solves a variational inequality problem defined by an affine mapping with an asymmetric matrix by solving a sequence of variational inequality problems, each of which is defined by an affine mapping with a symmetric matrix.

The following theorem shows that the iterates generated by the algorithm converge to the unique solution x^* if $\| S^{-1}M^T\|_S < 1$, where $S = M + M^T$.

Theorem 4.3

Let $f(x) = Mx - b$, where M is an nxn positive definite matrix; let $S = M + M^T$; and let C be a closed, convex, nonempty subset of R^n. Then, if $\| S^{-1}M^T\|_S < 1$, the sequence of iterates generated by the algorithm converges to the solution x^* of VI(f,C).

The proof of the theorem has a simple geometrical interpretation. Before proceeding with the details of the proof, let us briefly highlight the geometry underlying the argument. Let $c = \| S^{-1}M^T\|_S$. We will show that if the distance with respect to the S norm from the solution x^* to a point $\bar{x}\varepsilon C$ is greater than $c\| x^k - x^*\|_S$, then $(x^* - \bar{x})^T g(\bar{x}, x^k) < 0$; that is, the mapping $g(\bar{x}, x^k)$ points away from the point $x^*\varepsilon C$. This implies that

$\bar{x}$ cannot solve the k^{th} subproblem, since the subproblem solution x^{k+1} must satisfy $(x - x^{k+1})^T g(x^{k+1}, x^k) \geq 0$ for every $x \varepsilon C$. Therefore, the distance with respect to the S norm from x^* to x^{k+1} must be less than $c\| x^k - x^* \|_S$, which ensures, since $c < 1$, that the iterates contract to the solution in S norm. Figure 4.2 illustrates this geometrical idea. (The general structure of this proof is similar to that of Ahn's [1979] proof of the nonlinear Jacobi method.)

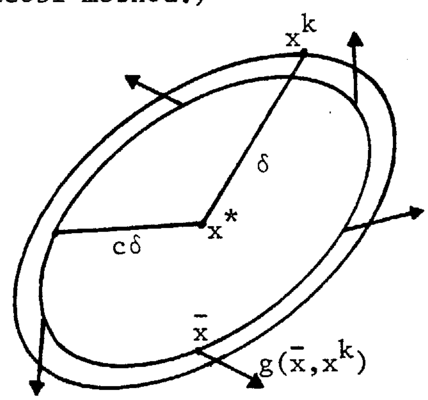


Figure 4.2

The Approximate Map $g(\bar{x},x^k)$ Points Away from x^* if $\| \bar{x}-x^* \|_S > c\|x^k-x^*\|_S$

Proof of Theorem 4.3

We show that

$$\| x^{k+1} - x^* \|_S \leq c\| x^k - x^* \|_S,$$

and, hence, that

$$\| x^{k+1} - x^* \|_S \leq (c)^{k+1} \| x^0 - x^* \|_S .$$

Because $c \varepsilon [0,1)$, the righthand side of this inequality approaches zero as $k \to \infty$, and, therefore, $\lim_{k\to\infty} x^k = x^*$.

Recall that x^{k+1} solves the subproblem

$$(x - x^{k+1})^T g(x^{k+1}, x^k) \geq 0 \quad \text{for every } x \epsilon C.$$

In particular, this inequality is valid for $x = x^*$.

Let $\delta = \| x^k - x^* \|$. Assume that $x^k \neq x^*$, and, hence, that $\delta > 0$. Let $\bar{x} \neq x^*$ be a point in C and let $\bar{c}$ be defined by $\| \bar{x} - x^* \|_S = \bar{c}\delta$. Then,

$$(x^* - \bar{x})^T g(\bar{x}, x^k)$$

$$= (x^* - \bar{x})^T[(M + M^T)\bar{x} - M^T x^k - b]$$

$$= (x^* - \bar{x})^T f(x^*) + (x^* - \bar{x})^T[(M + M^T)(\bar{x} - x^*) + M^T(x^* - x^k)]$$

$$< (x^* - \bar{x})(M + M^T)(\bar{x} - x^*) + (x^* - \bar{x})^T SS^{-1}M^T(x^* - x^k) \quad \text{since } \bar{x} \neq x^*$$

$$= -\| x^* - \bar{x} \|_S^2 + (x^* - \bar{x}, S^{-1}M^T(x^* - x^k))_S$$

$$\leq -\bar{c}^2 \delta^2 + \| x^* - \bar{x} \|_S \; \| S^{-1}M^T(x^* - x^k) \|_S \quad \text{by Cauchy's Inequality}$$

$$\leq -\bar{c}^2 \delta^2 + \bar{c}\, \delta \; \| S^{-1}M^T \|_S \; \| x^* - x^k \|_S$$

$$= -\bar{c}^2 \delta^2 + \bar{c}\, \delta \; c\, \delta$$

$$= \bar{c}\delta^2 (c - \bar{c}). \tag{4.6}$$

Now if $\bar{x} = x^{k+1}$, then, as noted above, $(x^* - \bar{x})^T g(\bar{x}, x^k) \geq 0$.

Therefore, $\bar{c} < c$, since by (4.6), $\bar{c}\delta^2(c - \bar{c}) > (x^* - \bar{x})^T g(\bar{x}, x^k) \geq 0$.

But then,

$$\| x^{k+1} - x^* \|_S = \bar{c}\| x^k - x^* \|_S < c\| x^k - x^* \|_S . \quad \blacksquare$$

The following example considers a constrained variational inequality problem defined by the same affine map as that in Example 4.1.

Example 4.2

Let $M = \begin{bmatrix} 1 & 2 \\ -2 & 4 \end{bmatrix}$ and $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and

let $C = \{x \varepsilon R^2: x_1 \geq 0,\ x_2 \geq 0, \text{ and } x_2 \leq (1/6)x_1 + 1/8\}$. The solution x^{k+1} to the k^{th} subproblem must satisfy

$$(x - x^{k+1})^T(Sx^{k+1} - M^T x^k - b) \geq 0 \qquad \text{for every } x \varepsilon C,$$

where $S = M + M^T = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$. Let $x^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then $x^1 = \begin{pmatrix} 21/20 \\ 3/10 \end{pmatrix}$,

$x^2 = \begin{pmatrix} 9/10 \\ 11/40 \end{pmatrix}$, $x^3 = \begin{pmatrix} 33/40 \\ 21/80 \end{pmatrix}$, $x^4 = \begin{pmatrix} 63/80 \\ 41/160 \end{pmatrix}$, ... The sequence $\{x^k\}$ converges to the solution $x^* = \begin{pmatrix} 3/4 \\ 1/4 \end{pmatrix}$ because $\| S^{-1}M^T \|_2 = \sqrt{2}/2 < 1$. Figure 4.3 illustrates the sequence of iterates $\{x^k\}$ as well as a sequence of ellipses.

We can also interpret the algorithm applied to constrained problems in terms of a sequence of ellipsoids. Given an iterate x^k, the algorithm selects the center of the ellipsoid E_0^k as the next iterate x^{k+1} if the center is a feasible point. Otherwise, we can determine x^{k+1} by finding the smallest ellipsoid about the center of E_0^k that contains a feasible point. This

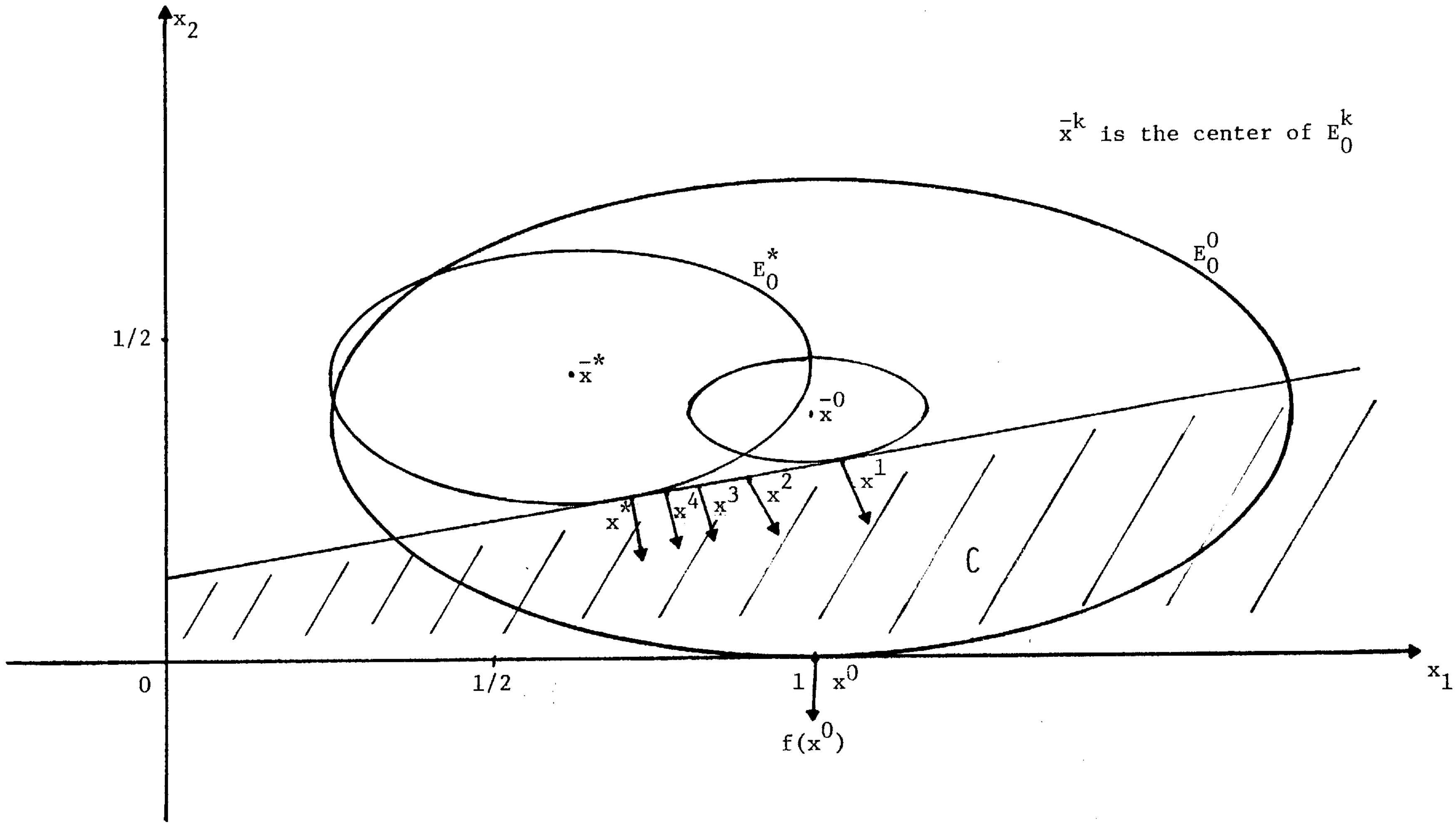


Figure 4.3 The Contracting Ellipsoid Method Solves a Constrained Affine Variational Inequality Problem

feasible point is the next iterate, x^{k+1}. This sequence of ellipses does not necessarily converge to the solution x^*. In fact, the point $x^* = \begin{pmatrix} 3/4 \\ 1/4 \end{pmatrix}$ determines a set of ellipsoids centered about the point $\begin{pmatrix} 5/8 \\ 7/16 \end{pmatrix}$. The smallest ellipse in this set containing a feasible point contains the point x^*, thus establishing that x^* is the solution to the problem.

4.1.3 Constrained Problems with Nonlinear Maps

In this subsection we consider the constrained variational inequality problem VI(f,C), where C is a closed, convex subset of R^n and f is a strictly monotone nonlinear map.

For this general problem, the objective function of the minimization subproblem

$$\min_{x \in C} (x - x^k)^T f(x) \tag{4.7}$$

is not necessarily convex. A solution to this subproblem must satisfy the first-order optimality conditions for problem (4.7); that is, x^{k+1} must satisfy

$$(x - x^{k+1})^T [\nabla^T f(x^{k+1})(x^{k+1} - x^k) + f(x^{k+1})] \geq 0 \quad \text{for every } x \in C.$$

In general, the mapping defining this variational inequality subproblem is neither monotone nor affine. To avoid solving this potentially difficult subproblem, for k = 0,1,... we modify the contracting ellipsoid algorithm at the k^{th} iteration by linearly approximating f about x^k. That is, by replacing f(x) in problem 4.7 with $f(x^k) + \nabla f(x^k)(x - x^k)$, we obtain the following algorithm.

General Contracting Ellipsoid Algorithm

Step 0: Select $x^o \varepsilon C$. Set $k = 0$.

Step 1: Let $x^{k+1} = \underset{x\varepsilon C}{\text{argmin}} \, [(x - x^k)^T f(x^k) + (x - x^k)^T \nabla f(x^k)(x - x^k)]$.

If $x^{k+1} = x^k$, then stop: $x^k = x^*$.

Otherwise, return to Step 1 with $k = k+1$. ∎

The strict monotonicity of f ensures that the k^{th} subproblem is a strictly convex quadratic programming problem. The unique solution x^{k+1} to this subproblem must, therefore, satisfy the necessary and sufficient optimality conditions.

$$(x - x^{k+1})^T[[(\nabla f(x^k) + \nabla^T f(x^k)] \; (x^{k+1} - x^k) + f(x^k)] \geq 0 \tag{4.8}$$

for every $x \varepsilon C$.

Let $M_k = \nabla f(x^k)$ and let $g(x, x^k)$ be the mapping defining the variational inequality subproblem on the k^{th} iteration; i.e.,

$$g(x, x^k) = (M_k + M_k^T)(x - x^k) + f(x^k). \tag{4.9}$$

The general contracting ellipsoid algorithm, therefore, solves a nonlinear variational inequality problem by solving a sequence of variational inequality problems, each of which is defined by an affine mapping with a symmetric matrix.

The following theorem establishes convergence conditions for the general algorithm. The general structure of the proof of the theorem is similar to that of the proof of Theorem 4.3. Here, we show that if $\| S^{-1}M^T\|_S < 1$, where

$M = \nabla f(x^*)$ and $S = M + M^T$, then there exists a constant $r\varepsilon[0,1)$ such that $\| x^{k+1} - x^* \|_S \leq r \| x^k - x^* \|_S$. To do this, we again show that if the distance with respect to the S norm from the solution x^* to point $\bar{x}\varepsilon C$ is greater than $r \| x^k - x^* \|_S$, then the mapping $g(\bar{x}, x^k)$ points away from the solution x^*, which shows that $\bar{x}$ cannot solve the variational inequality subproblem. The convergence proof requires f to be twice Gateaux differentiable in order to use the second derivative of f to bound the error in making a linear approximation to f. The theorem also assumes that a solution x^* to VI(f,C) exists. This assumption is necessary because we do not assume that f is uniformly monotone or that C is compact.

Theorem 4.4

Let f be strictly monotone and twice Gateaux differentiable, and let C be a closed convex subset of R^n. Assume that a solution x^* to VI(f,C) exists, and that $\| S^{-1}M^T \|_S < 1$, where $M = \nabla f(x^*)$ and $S = M + M^T$. Then, if the initial iterate x^0 is sufficiently close in S norm to the solution x^*, the sequence of iterates generated by the general contracting ellipsoid algorithm contracts to the solution in M norm.

Proof

We show that there exists an $r\varepsilon[0,1)$ satisfying

$$\|x^{k+1} - x^*\|_S \leq r \| x^k - x^* \|_S,$$

and, hence, that

$$\|x^{k+1} - x^*\|_S \leq (r)^{k+1} \| x^0 - x^* \|_S.$$

Because $r \varepsilon [0,1)$, $\lim_{k\to\infty} (r)^{k+1} = 0$, and, hence, $\lim_{k\to\infty} x^k = x^*$,

Let $c = \| S^{-1}M^T \|_S$, and

Let $K = \sup_{\|x-x^*\| \le 1} \left\{ \sup_{0 \le t \le 1} \| S^{-1}[\nabla^2 f(x + t(x^* - x)] \|_S \right\}$.

Two extended mean value theorems (3.3.5 and 3.3.6 in Ortega and Rheinboldt [1970]) show that,

if $\| x - x^* \| \le 1$, then

$$\| S^{-1}[f(x) - (f(x^*) + \nabla f(x^*)(x - x^*))] \|_S \le K \| x - x^* \|_S^2 , \tag{4.10}$$

and

$$\| S^{-1}[\nabla f(x) - \nabla f(x^*)] \|_S \le K \| x - x^* \|_S . \tag{4.11}$$

Let $\delta = \| x^k - x^* \|_S$, and let $\gamma > 0$ satisfy $\gamma < \mathrm{Min}(\frac{1-c}{5K}, 1)$. (Note that $\frac{1-c}{5K} > 0$, since $c < 1$.) Assume that $0 < \delta < \gamma$. Finally, let $r = (c + 3K\gamma/1-2K\gamma)$. By definition of γ, $r < 1$.

Recall that x^{k+1} solves the subproblem (4.8):

$$(x - x^{k+1})^T g(x^{k+1}, x^k) \ge 0 \qquad \text{for every } x \varepsilon C.$$

In particular, this inequality is valid for $x = x^*$. Let $\bar{x} \ne x^*$ be a point in c and let $\bar{c}$ be defined by $\| \bar{x} - x^* \|_S = \bar{c}\delta$. Then, we obtain the following chain of inequalities:

$$(x^* - \bar{x})^T g(\bar{x}, x^k)$$

$$= (x^* - \bar{x})^T [(M_k + M_k^T)(\bar{x} - x^k) + f(x^k)]$$

$$= (x^* - \bar{x})^T f(x^*) + (x^* - \bar{x})^T [(M + M^T)(\bar{x} - x^*) + M^T(x^* - x^k)$$

$$+ M(x^* - x^k) + f(x^k) - f(x^*) + (M_k - M)(\bar{x} - x^k)$$

$$+ (M_k^T - M^T)(\bar{x} - x^k)]$$

$$< (x^* - \bar{x})^T S(\bar{x} - x^*) + (x^* - \bar{x})^T SS^{-1} M^T(x^* - x^k)$$

$$+ (x^* - \bar{x}) SS^{-1} [M(x^* - x^k) + f(x^k) - f(x^*)]$$

$$+ (x^* - \bar{x})^T SS^{-1}(M_k - M)(\bar{x} - x^k)$$

$$+ (\bar{x} - x^k)^T SS^{-1}(M_k - M)(x^* - \bar{x}),$$

where the second equality is a result of adding and subtracting terms so that we can obtain expressions in terms of the S norm, and the inequality holds because $(\bar{x} - x^*)^T f(x^*) > 0$.

We consider each of the terms in the last expression separately:

$$(x^* - \bar{x})^T S(\bar{x} - x^*) = -\| \bar{x} - x^* \|_S^2 = -(\bar{c}\delta)^2;$$

$$(x^* - \bar{x})^T S S^{-1} M^T (x^* - x^k) =$$

$$= (x^* - \bar{x}, S^{-1} M^T (x^* - x^k))_S$$

$$\leq \| x^* - \bar{x} \|_S \ \| S^{-1} M^T (x^* - x^k) \|_S \qquad \text{by Cauchy's Inequality}$$

$$\leq \| x^* - \bar{x} \|_S \ \| S^{-1} M^T \|_S \ \| x^* - x^k \|_S$$

$$= \bar{c}\ \delta\ c\ \delta;$$

$$(x^* - \bar{x}) S S^{-1} [f(x^k) - f(x^*) - M(x^k - x^*)]$$

$$\leq \| x^* - \bar{x} \|_S \ \| S^{-1} [f(x^k) - f(x^*) - M(x^k - x^*)] \|_S$$

$$\leq \| x^* - \bar{x} \|_S K \| x^* - x^k \|_S^2 \qquad \text{as above}$$

$$= \bar{c}\ \delta\ K\delta^2; \qquad \text{by (4.10)}$$

$$(x^* - \bar{x})^T S S^{-1} (M_k - M)(\bar{x} - x^k)$$

$$\leq \| x^* - \bar{x} \|_S \ \| S^{-1} (M_k - M) \|_S \ \| \bar{x} - x^k \|_S \qquad \text{as above}$$

$$\leq \| x^* - \bar{x} \|_S \ K \| x^k - x^* \|_S \ \| \bar{x} - x^k \|_S \qquad \text{by (4.11)}$$

$$\leq \| x^* - \bar{x} \|_S \ K \| x^k - x^* \|_S \ (\| \bar{x} - x^* \|_S + \| x^* - x^k \|_S)$$

by the triangle inequality

$$= \bar{c}\ \delta\ K\delta(\bar{c}\delta + \delta);$$

and, similarly, $(\bar{x} - x^k)^T SS^{-1}(M_k - M)(x^* - \bar{x}) \leq (\bar{c}\delta + \delta)K\delta\bar{c}\delta$.

Combining the above inequalities, we obtain

$$(x^* - \bar{x})^T g(\bar{x}, x^k) < \bar{c}\delta^2(-\bar{c} + c + \delta K + K\bar{c}\delta + K\delta + K\bar{c}\delta + K\delta)$$

$$= \bar{c}\delta^2(c + 3K\delta - \bar{c}(1 - 2K\delta))$$

$$< \bar{c}\gamma^2(c + 3K\gamma - \bar{c}(1 - 2K\gamma)). \tag{4.12}$$

Now if $\bar{x} = x^{k+1}$, then, as noted above $(x^* - \bar{x})^T g(\bar{x}, x^k) \geq 0$. Therefore,

$r = \dfrac{3 + K\gamma}{1-2K\gamma} > \bar{c}$, since by (4.12), $\bar{c}\gamma^2(c + 3K\gamma - \bar{c}(1-2K\gamma)) >$

$(x^* - \bar{x})^T g(\bar{x}, x^k) \geq 0$.

But then,

$$\| x^{k+1} - x^* \|_S = \bar{c}\| x^k - x^* \|_S < r\| x^k - x^* \|_S .$$ ∎

4.1.4 Further Geometrical Considerations

In this subsection, we interpret the steepest descent method in terms of the ellipsoidal level sets $\{E_0^k\}$ that are intrinsic in the contracting ellipsoid method, and discuss the role of symmetry in the concentric ellipsoid algorithm. We also compare convergence conditions for these two methods.

Let $VI(f,R^n)$ be an unconstrained variational inequality problem defined by the affine map f. Consider the sequence of ellipses

$$E_0^k = \{x:(x - x^k)^T f(x) \leq 0\}.$$

By definition of E_0^k, the chord from x^k to any point $x \in \partial E_0^k$ is orthogonal to the vector $f(x)$. In addition, the vector $f(x^k)$ is normal to the tangent plane of E_0^k at x^k. Recall that, given x^k, the steepest descent method determines the next iterate

$$x^{k+1} = x^k - \theta_k f(x^k),$$

where θ_k is chosen so that $f^T(x^{k+1})f(x^k) = 0$. Thus, on the k^{th} iteration, the algorithm moves from x^k in the $-f(x^k)$ direction to the point x^{k+1} at which $f(x^{k+1})$ is orthogonal to the direction of movement $x^{k+1} - x^k$. In terms of the ellipsoid E_0^k, the steepest descent method moves from x^k to the point on ∂E_0^k that is in the $-f(x^k)$ direction, since at that point $f(x^{k+1})$ is orthogonal to the direction of movement $x^{k+1} - x^k$. Figure 4.4 illustrates this interpretation of the steepest descent direction. In the figure, x_{SD}^{k+1} denotes the iterate obtained from the point x^k by the generalized steepest descent direction, while x_{CE}^{k+1} denotes the iterate obtained from the point x^k by the contracting ellipsoid method. Note that the steepest descent direction is not, in general, the same as the direction of movement given by the contracting ellipsoid method.

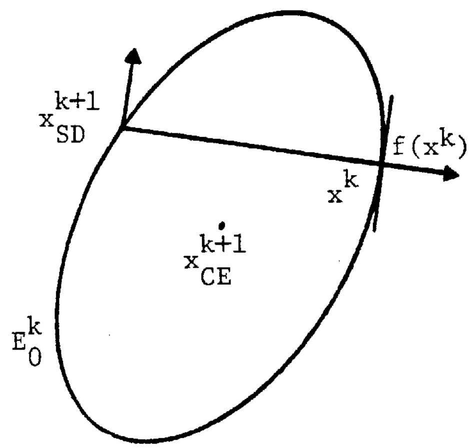


Figure 4.4: Relationship Between the Ellipsoid Level Sets and the Steepest Descent Method

When $f(x) = Mx - b$, and M is symmetric and positive definite, the contracting ellipsoid method generates the sequence $\{x^k\}$ for the unconstrained problem VI(f,C) by

$$x^{k+1} = (2M)^{-1}(Mx^k + b)$$

$$= (1/2)(x^k - x^*), \quad k = 0,1,\ldots$$

Hence, the algorithm moves halfway to the solution on each iteration. In this case, the ellipsoids E_0^k are tangent to each other at x^*, as illustrated in Figure 4.5. Note that even if M was the identity matrix, the algorithm would still move halfway to the solution. In this instance, the steepest descent algorithm would converge to x^* in a single iteration. In general, though, we expect that the contracting ellipsoid algorithm would outperform the steepest descent algorithm. (See the discussion in Chapter 5)

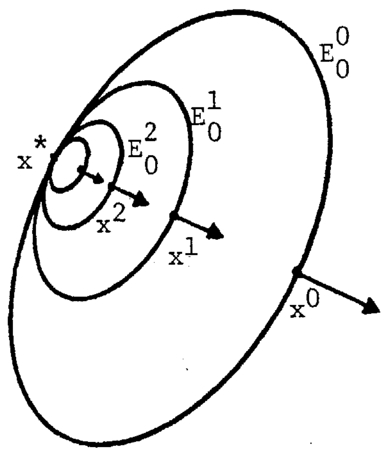


Figure 4.5: The Contracting Ellipsoid Iterates Move Halfway to the Solution if M is Symmetric

If the positive definite matrix M is not symmetric, both the steepest descent method and the contracting ellipsoid method are guaranteed to converge only if some restriction on the degree of asymmetry of M is imposed. For unconstrained problems, the generalized steepest descent algorithm converges if and only if M^2 is positive definite, while the contracting ellipsoid algorithm converges if and only if $\rho(S^{-1}M^T) < 1$, where $S = M + M^T$. For constrained problems, the contracting ellipsoid method is guaranteed to converge if $\| S^{-1}M^T \|_S < 1$. Table 4.1 compares these conditions for 2x2 matrices. Recall that if M^2 is positive definite, then $\rho(S^{-1}M^T) < 1$, and that $\| S^{-1}M^T \|_S < 1$ implies that $\rho(S^{-1}M^T) < 1$. Although in the 2x2 case, the conditions $\rho(S^{-1}M^T) < 1$ and $\| S^{-1}M^T \|_S < 1$ are identical, we do not expect this to be true in general.

	$M = \begin{bmatrix} 1 & r \\ -r & 1 \end{bmatrix}$	$M = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$	$M = \begin{bmatrix} 1 & b \\ c & 1 \end{bmatrix}$	$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$
M positive definite	all values of r	$\lvert b \rvert < \sqrt{4ad}$	$\lvert b + c \rvert < 2$	$(b + c)^2 < 4ad$
M^2 positive definite	$\lvert r \rvert < 1$	$\lvert b \rvert < \frac{2ad}{2+d}$ $< 2 \min(a,d)$	$b^2+c^2<1+b^2c^2$	$(a+b)^2(b+c)^2$ $-4bc(a^2+d^2)$ $<4a^2d^2+4b^2c^2$
$\rho(S^{-1}M^T) < 1$	$\lvert r \rvert < \sqrt{3}$	$\lvert b \rvert < \sqrt{3ad}$	$b^2+c^2<3-bc$	b^2+bc+c^2 $<3ad$
$\lVert S^{-1}M^T \rVert_S < 1$	$\lvert r \rvert < \sqrt{3}$	$\lvert b \rvert < \sqrt{3ad}$	$b^2+c^2<3-bc$	b^2+bc+c^2 $<3ad$

Table 4.1: A Comparison of Convergence Condition for 2x2 Matrices

4.1.5 The Relationship Between the Contracting Ellipsoid Method and Other Algorithms

The contracting ellipsoid algorithm is closely related to several algorithms for solving systems of equations and variational inequality problems. In this subsection, we discuss its relationship to matrix splitting algorithms, projection algorithms, and a general iterative algorithm devised by Dafermos [1983]. In section 4.2, we discuss the subgradient algorithm for solving a max-min problem that is equivalent to the problrm VI(f,C), and show that it iteratively solves the same subproblem as the contracting ellipsoid method.

Recall that the contracting ellipsoid method solves $VI(f,R^n)$, where $f(x) = Mx - b$, by iteratively solving the recursion (4.3):

$$x^{k+1} = S^{-1}(M^T x^k + b),$$

where $S = M + M^T$. For the unconstrained problem defined by an affine map, this algorithm is a special case of a general iterative method to solve linear equations based on the principle of matrix splitting. For a linear system $Mx = b$, splitting the matrix M into the sum

$$M = A - B,$$

where A is nonsingular, produces an equivalent linear system

$$Ax = Bx + b,$$

or, equivalently,

$$x = A^{-1}(Bx + b).$$

This matrix splitting induces an iterative method defined by

$$x^{k+1} = A^{-1}(Bx^k + b)$$

$$= x^k - A^{-1}(Mx^k - b), \quad k = 0,1,2,\ldots$$

The Jacobi, Gauss-Seidel, and successive overrelaxation methods are examples of iterative methods induced by matrix splittings. The matrix splitting

$$M = (M + M^T) - M^T$$

induces the recursion (4.3) that defines the contracting ellipsoid method.

The contracting ellipsoid method solves VI(f,C), where $f(x) = Mx - b$, by iteratively solving the subproblem (4.5):

$$\min_{x \in C} (x - x^k)^T f(x)$$

The following lemma shows that the subproblem (4.5) is a projection step, with a steplength of one. Hence, the contracting ellipsoid method for problems with affine maps is a projection method, with the metric of the projection defined by the S norm and the steplength at each iteration equal to one. (If f is nonlinear, the subproblem of the general contracting ellipsoid method is also a projection step with a steplength of one.)

Lemma 4.2

If $f(x) = Mx - b$ and M is positive definite, then the subproblem

$$x^{k+1} = \underset{x \varepsilon C}{\operatorname{argmin}}\ (x - x^k)^T f(x)$$

is equivalent to the projection

$$x^{k+1} = P_C^S[x^k - S^{-1} f(x^k)]$$

defined by the matrix $S = M + M^T$ and the projection operator P_C^S onto the set C with respect to the S norm.

Proof

$x^{k+1} = P_C^S[x^k - S^{-1}f(x^k)]$ if and only if x^{k+1} is the point in C that is closest to $x^k - S^{-1}f(x^k)$ in S norm; i.e., if and only if

$$x^{k+1} = \underset{x \varepsilon C}{\operatorname{argmin}}\ \| x^k - S^{-1}f(x^k) - x\|_S^2$$

$$= \underset{x \varepsilon C}{\operatorname{argmin}}\ (x^k - S^{-1}f(x^k) - x)^T S(x^k - S^{-1}f(x^k)-x)$$

$$= \underset{x \varepsilon C}{\operatorname{argmin}}\ (Sx^k - f(x^k) - Sx)^T S^{-1}(Sx^k - f(x^k) - Sx)$$

$$= \underset{x \varepsilon C}{\operatorname{argmin}}\ [M^T x^k + b - (M + M^T)x]^T S^{-1}[M^T x^k + b - (M + M^T)x]$$

$$= \underset{x \varepsilon C}{\operatorname{argmin}}\{(M^T x^k + b)^T S^{-1}(M^T x^k + b) - 2x^T(M^T x^k + b) + x^T Sx\}$$

$$= \underset{x \varepsilon C}{\operatorname{argmin}}\ 2[x^T Mx - x^T(M^T x^k + b)] \qquad \text{(dropping the constant term)}$$

$$= \underset{x \in C}{\operatorname{argmin}} \{x^T(Mx - b) + x^T M^T x^k\}$$

$$= \underset{x \in C}{\operatorname{argmin}} \ (x - x^k)^T(Mx - b) \qquad \text{(adding a constant term)}$$

$$= \underset{x \in C}{\operatorname{argmin}} \ (x - x^k)^T f(x). \qquad \blacksquare$$

Finally, the contracting ellipsoid method for problems defined by affine maps fit into the framework of the general iterative scheme devised by Dafermos [1983]. (See Section 2.2.2 for a more detailed description of her general algorithm.) The general scheme solves VI(f,C) by constructing a mapping g(x,y) that approximates the mapping f(x) about the point y so that

(i) $g(x,x) = f(x)$ for every $x \in C$; and

(ii) $g_x(x,y)$ is symmetric and positive definite for every $x \in C$, and $y \in C$, where $g_x(x,y)$ denotes the derivative of g with respect to the first component.

The algorithm converges if g satisfies

$$\| [g_x^{1/2}(x_1,y_1)]^{-T} g_y(x_2,y_2)[g_x^{1/2}(x_3,y_3)]^{-1} \|_2 < 1 \tag{4.13}$$

for every $x_1,x_2,x_3,y_1,y_2,y_3 \in C$.

Because the contracting ellipsoid method iteratively determines the point $x^{k+1} \in C$ satisfying

$$(x-x^{k+1})^T[(M + M^T)x^{k+1} - (M^T x^k + b)] \geq 0 \quad \text{for every } x \in C,$$

the algorithm fits into the general scheme described above with

$$g(x,y) = (M + M^T)x - (M^T y + b).$$

Conditions (i) and (ii) are satisfied because

$$g(x, x) = Mx - b = f(x)$$

and

$$g_x(x, y) = M + M^T \text{ is positive definite and symmetric.}$$

Because $g_y(x,y) = -M^T$, the conditions (4.13) reduce to

$$\| (S^{\frac{1}{2}})^{-T} M^T (S^{\frac{1}{2}})^{-1} \|_2 < 1.$$

Thus, since $\| (S^{\frac{1}{2}})^{-T} M^T (S^{\frac{1}{2}})^{-1} \|_2 = \| (S^{\frac{1}{2}})^{-1} (S^{\frac{1}{2}})^{-T} M^T (S^{\frac{1}{2}}) \|_S = \| S^{-1} M^T \|_S$, the conditions (4.10) reduce to the sufficient condition for convergence specified in Theorem 4.3.

When f is not affine, the mapping $g(x,x^k)$ defining the variational inequality subproblem of the contracting ellipsoid method; that is, the mapping

$$g(x,x^k) = \nabla^T f(x)(x-x^k) + f(x),$$

is not necessarily monotone in x (as required by condition (ii) stated above.) The algorithm for a problem defined by a nonlinear map does not, therefore, fit into Dafermos' general framework. The modification of the contracting ellipsoid algorithm that we discussed in section 4.1.3 does, however, fit into this framework because

$$g(x,y) = [\nabla f(y) + \nabla^T f(y)](x-y) + f(y)$$

satisfies

(i) $g(x,x) = f(x)$

and

(ii) $g_x(x,y) = \nabla f(y) + \nabla^T f(y)$ is positive definite and symmetric for every $y \in C$ because f is strictly monotone.

The conditions for convergence (4.13) are

$$\| ([\nabla f(y_1) + \nabla^T f(y_1)]^{\frac{1}{2}})^{-T} \{[\nabla^2 f(y_2) + (\nabla^2 f(y_2))^T](x_2 - y_2) - f^T(y_2)\} \cdot ([\nabla f(y_3) + \nabla^T f(y_3)]^{\frac{1}{2}})^{-1} \|_2 < 1 .$$

These conditions are clearly much more difficult to verify than those specified in Theorem; namely,

$$\|S^{-1} M^T\|_S < 1,$$

where $M = \nabla f(x^*)$ and $S = M + M^T$.

4.2 Subgradient Algorithms

In this section, we discuss a class of subgradient algorithms that can be used to maximize nondifferentiable, concave functions. Shor [1964] originally suggested using gradient methods to solve nondifferentiable optimization problems. Polyak [1967] obtained results on constrained problems that extended Shor's earlier results on unconstrained problems.

Consider the maximization problem

$$\max_{x \in C} F(x),$$

where F is a nondifferentiable concave, continuous function and C is a closed convex subset of R^n. Given the previous iterate x^k, a subgradient algorithm determines a subgradient of F at x^k; i.e., a vector $\gamma_k \in R^n$ satisfying

$$F(x) \leq F(x^k) + \gamma_k(x - x^k) \quad \text{for every } x \in R^n.$$

and a steplength α_k. It then generates the $(k+1)^{st}$ iterate by

$$x^{k+1} = P_C[x^k - \alpha_k\gamma_k],$$

where P_c is the projection operator onto the set C. Polyak [1969] proposes the use of a steplength α_k given by

$$\alpha_k = \lambda_k \frac{F(x^*) - F(x^k)}{\|\gamma_x\|^2},$$

where $0 < \varepsilon_1 \leq \lambda_k \leq 2 - \varepsilon_2 < 2$ and x^* maximizes F over C. He discusses several methods for choosing α_k, and analyzes the convergence properties of the subgradient algorithm. (In general, the rate of convergence of the algorithm can be rather slow.)

Consider the constrained variational inequality problem VI(f,C). Assume that the problem is formulated over a closed, convex ground set $C \subseteq R^n$, and that the mapping $f:R^n \to R^n$ is monotone and continuously differentiable. Recall from Section 2.4 that under these assumptions, the problem VI(f,C) is equivalent to the max-min problem

$$\underset{x \in C}{\text{Max}} \left\{ \underset{y \in C}{\text{Min}} \ (y-x)^T f(y) \right\}. \tag{4.14}$$

If we define

$$H(x) := \underset{y \in C}{\text{Min}} (y-x)^T f(y),$$

then problem (4.14) can be restated as the nonlinear maximization problem

$$\underset{x \in C}{\text{Max}} \ H(x). \tag{4.15}$$

Because H(x) is the pointwise minimum of functions $(y-x)^T f(y)$ that are linear in x, H(x) is concave. Problem (4.15) is, therefore, a convex programming problem. Clearly, $H(x) \leq 0$ for every $x \in C$; moreover, $H(x^*) = 0$ if and only if x^* solves VI(f,C).

The reformulation of VI(f,C) as the max-min problem (4.14) (or, equivalently, (4.15)) motivates a number of algorithms to solve VI(f,C). For example, Auslender [1976] and Nguyen and Dupuis [1981] devise algorithms that approximate H(x) on the k^{th} iteration by the piecewise linear function

$$H_k(x) := \text{Min}\{(x^i - x)^T f(x^i) : i = 0,1,\ldots,k\}.$$

These algorithms assume that either f is uniformly monotone or that f is strictly monotone and C is compact.

The max-min formulation also suggests using a subgradient algorithm to solve VI(f,C). The functional $H(x) := \min_{y \in C} (y-x)^T f(y)$ is concave, and, in general, nondifferentiable. Thus, the subgradient algorithm can be applied to problem (4.15). Note that we need not assume that f is strictly or uniformly monotone on C. f must be monotone, however, so that VI(f,C) can be reformulated as the max-min problem (4.14).

Let $\partial H(x)$ denote the subdifferential of H at the point x; that is, the set of subgradients of H at x. Because H(x) is the pointwise minimum of the functions $(y-x)^T f(y)$, $\partial H(x)$ is given by the convex hull of the gradients of those functions $(\bar{y} - x)^T f(\bar{y})$ for which $\bar{y} = \text{argmin}\{(y-x)^T f(y): y \in C\}$. Therefore, $\partial H(x)$ is given by

$$\partial H(x) = \{-f(\bar{y}): \bar{y} \in \underset{y \in C}{\text{argmin}}\ (y-x)^T f(y)\}.$$

For most problems, the value of the function at the optimal solution must be estimated at each iteration in order to specify the steplength at that iteration. For problem (4.15), however, the expression for the steplength α_n can be simplified, because the value of H at the solution x^* is zero. Thus, the subgradient algorithm to solve

$$\underset{x \in C}{\text{Min}}\ H(x)$$

can be stated as follows:

Subgradient Algorithm for VI(f,C)

Step 0: Select $x^0 \varepsilon C$. Set $k = 0$.

Step 1: Let $x^{k+1} = P_C[x^k - \alpha_k f(\bar{x}^k)]$,

where $\bar{x}^k = \text{argmin } \{(x-x^k)^T f(x) : x \varepsilon C\}$

$$\text{and } \alpha_k = \frac{-\lambda_k H(x^k)}{\| f(\bar{x}^k) \|^2} = -\lambda_k \cdot \frac{(\bar{x}^k - x^k)^T f(\bar{x}^k)}{\| f(\bar{x}^k) \|^2},$$

and $0 < \varepsilon_1 \leq \lambda_k \leq 2 - \varepsilon_2 \leq 2$.

If $x^{k+1} = x^k$, Stop: $x^k = x^*$.

Otherwise, return to Step 1 with $k = k+1$. ∎

(Note that the subproblem that determines $\bar{x}^k$ is exactly the same subproblem that determines x^{k+1} in the contracting ellipse algorithm. The subgradient algorithm does not move to the point $\bar{x}^k$; instead, it moves in the $-f(\bar{x}^k)$ direction from x^k.)

The idea of solving VI(f,C) by moving in the direction $-f(\bar{x}^k)$ from x^k, where

$$\bar{x}^k = \underset{x \varepsilon C}{\text{argmin}} \ (x - x^k)^T f(x),$$

is reminiscent of the "extragradient" algorithm proposed by Korpelevich [1977]. This modified projection algorithm will solve variational inequality problems defined by monotone mappings. (The usual projection algorithm, described in Chapter 2, requires f to be uniformly monotone). The extragradient method moves in the direction $-f(\tilde{x}^k)$ from x^k, where

$$\tilde{x}^k = P_C[x^k - \alpha f(x^k)].$$

The algorithm can be stated as follows:

<u>Extragradient Algorithm</u>

Step 0: Select $x^0 \varepsilon C$. Set $k=0$.

Step 1: Let $\tilde{x}^k = P_C[x^k - \alpha f(x^k)]$.

If $\tilde{x}^k = x^k$, stop: $x^k = x^*$.

Otherwise, go to Step 2.

Step 2: Let $x^{k+1} = P_C[x^k - \alpha f(\tilde{x}^k)]$.

Go to Step 1 with $k = k+1$. $\blacksquare$

Korpelevich shows that the algorithm converges if the following conditions are satisfied:

(i) C is closed and convex;

(ii) f is monotone and Lipschitz continuous with Lipschitz coefficient L; and

(iii) the steplength $\alpha \varepsilon (0, \frac{1}{L})$.

The similarity between the subgradient and extragradient algorithms is more than superficial. Indeed, if $f(x) = Mx - b$, then recall from section 4.1.5 that the solution $\bar{x}^k$ to the k^{th} subproblem in the subgradient algorithm is a projection; in fact,

$$\bar{x}^k = P_C^S[x^k - S^{-1}f(x^k)],$$

where $S = M + M^T$.

4.3 The Frank-Wolfe Altorithm

In this section, we consider the constrained variational inequality problem VI(f,C)

Find $x^* \varepsilon C$ satisfying $(x - x^*)^T f(x^*) \geq 0$ for every $x \varepsilon C$,

where $f: C \subseteq R^n \to R^n$ is continuously differentiable and strictly monotone and C is a bounded polyhedron.

If $\nabla f(x)$ is symmetric for every x in C, then f is a gradient mapping; i.e., $f(x) = [\nabla F(x)]^T$ for some strictly convex functional $F: C \to R^1$. In this case, the unique solution x^* to the variational inequality problem solves the minimization problem

$$\operatorname*{Min}_{x \varepsilon C} F(x).$$

Thus, when f is a gradient mapping, the solution to the variational inequality problem may be found by using the Frank-Wolfe method to find the minimum of F over C.

The Frank-Wolfe algorithm is a linear approximation method that iteratively approximates F(x) by $F^k(x) + \nabla F(x^k)(x - x^k)$. On the k^{th} iteration, the algorithm determines a vertex solution v^k to the linear program

$$\operatorname*{Min}_{x \varepsilon C} F^k(x),$$

and then chooses as the next iterate the point x^{k+1} that minimizes F over the line segment $[x^k, v^k]$.

Frank-Wolfe Algorithm for Linearly Constrained Convex Minimization Problems

Step 0: Find $x^0 \varepsilon C$. Set $k = 0$.

Step 1: Given x^k, let v^k be a vertex solution to the linear program

$$\underset{x \varepsilon C}{\text{Min}}\ x^T \nabla F(x^k).$$

If $(v^k)^T \nabla F(x^k) = (x^k)^T\ F(x^k)$, then stop: $x^k = x^*$.

Otherwise, go to Step 2.

Step 2: Let w_k solve the one dimensional minimization problem:

$$\underset{0 \leq w \leq 1}{\text{Min}}\ F((1-w)x^k + wv^k)$$

Go to Step 1 with $x^{k+1} = (1-w_k)x^k + w_k v^k$ and $k = k+1$. ∎

This algorithm has been particularly effective in solving large-scale traffic equilibrium problems (see, for example, Bruynooghe et al. [1968], LeBlanc et al. [1975], and Golden [1975].) In this context, the linear program in Step 1 decomposes into a set of shortest path problems, one for each origin-destination pair. Therefore, the algorithm alternately solves shortest path problems (Step 1) and one-dimensional minimization problems (Step 2). (See section 1.4 for a discussion of the traffic equilibrium problem.)

The following theorem summarizes the convergence properties of the Frank-Wolfe algorithm. (See, for example, Martos [1975].)

Theorem 4.5

If F is pseudoconvex and continuously differentiable on C, and C is a bounded polyhedron, then the Frank-Wolfe algorithm produces a sequence $\{x^k\}$ for $k=1,2,\ldots$ of feasible points to the problem Min $\{F(x):x\varepsilon C\}$. The sequence is either finite, terminating with an optimal solution, or it is infinite, and has some accumulation points, any of which is an optimal solution.

When f is a gradient mapping over C, we can solve the linearly constrained variational inequality problem by reformulating the problem as the equivalent minimization problem and applying the Frank-Wolfe method. Equivalently, we can adapt the Frank-Wolfe method to solve the variational inequality problem directly.

Generalized Frank-Wolfe Method for the Linearly Constrained Variational Inequality Problem

Step 0: Find $x^0 \varepsilon C$. Set $k = 0$.

Step 1: Given x^k, let v^k be a vertex solution to the linear program

$$\min_{x\varepsilon C} x^T f(x^k).$$

If $(x^k)^T f(x^k) = (v^k)^T f(x^k)$, then stop: x^k is a solution to VI(f,C).

Otherwise, go to Step 2.

Step 2: Let w_k solve the following one-dimensional variational inequality problem on the line segment $[x^k, v^k]$:

Find $w_k \varepsilon [0,1]$ satsifying

$$\{[(1-w)x^k + wv^k] - [(1-w_k)x^k + w_k v^k]\}^T f[(1-w_k)x^k + w_k v^k] \geq 0$$

every $w \in [0,1]$.

Go to Step 1 with $x^{k+1} = (1-w_k)x^k + w_k v^k$ and $k = k+1$. ∎

Lemma 4.3

When $f(x) = \nabla f(x)$ for every x in R^n, the Frank-Wolfe algorithm for the linearly constrained minimization problem is equivalent to the Frank-Wolfe algorithm for the linearly constrained variational inequality problem.

Proof

The equivalence of Step 1 in the two algorithms is clear because $f(x) = \nabla F(x)$. Furthermore, the one-dimensional minimization problem of Step 2 in the first algorithm is equivalent to the one-dimensional variational inequality problem of Step 2 in the second algorithm; because F is convex, the variational inequality subproblem in the second algorithm defines the necessary and sufficient optimality conditions for the minimization subproblem in the first algorithm. ∎

If f is not a gradient mapping, the Frank-Wolfe algorithm need not converge to a solution of the variational inequality problem. The following two examples illustrate situations for which the sequence of iterates produced by the algorithms cycle among the extreme points of the feasible region. The first is a simple two dimensional example; the second could model delay time in a traffic equilibrium problem with one origin-destination pair and three parallel arcs. The mapping f is affine and strictly monotone in each of these examples.

Example 4.3

Let $f(x) = Mx - b$, where $M = \begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{bmatrix}$ and $b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, and let

$C = \{x = (y,z) : z \leq 1/2,\ z + \sqrt{3}\, y \geq -1,\ z - \sqrt{3}\, y \geq -1\}$.

The solution to $\overline{\text{VI}}(f,C)$ is $x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Let $x^0 = \begin{pmatrix} -\sqrt{3}/2 \\ 1/2 \end{pmatrix}$. The linear program of Step 1 of the generalized Frank-Wolfe algorithm solves at $v^0 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$, and the variational inequality subproblem of Step 2 solves at $x^1 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$. Continuing in this manner, the algorithm then generates $v^1 = \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix}$, $x^2 = \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix}$, $v^2 = \begin{pmatrix} -\sqrt{3}/2 \\ 1/2 \end{pmatrix}$, and $x^3 = \begin{pmatrix} -\sqrt{3}/2 \\ 1/2 \end{pmatrix} = x^0$. Hence, the iterates cycle about the three points x^0, x^1 and x^2. Figure 4.6 illustrates this cyclic behavior.

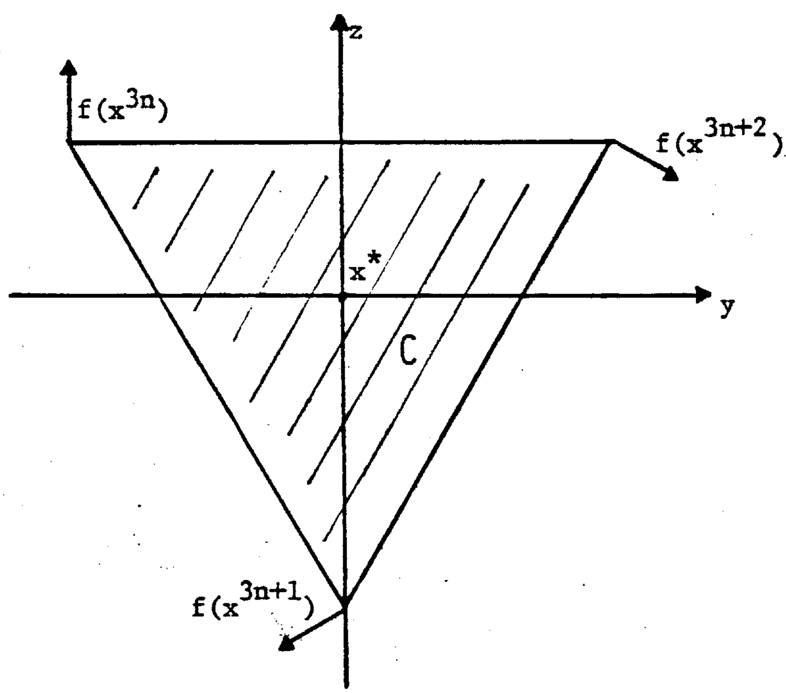


Figure 4.6 The Generalized Frank-Wolfe Algorithm Cycles

Example 4.4

Let $f(x) = Mx - b$, where $M = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$,.

and let $C = \{x=(x_1,x_2,x_3): x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0,\ x_1 + x_2 + x_2 = 1\}$.

The solution to VI(f,C) is $x^* = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$, since

$$(x - x^*)^T f(x^*) = (x_1-1/3,\ x_2-1/3,\ x_3-1/3) \begin{bmatrix} 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

$$= 2/3(x_1+x_2+x_3-1) = 0 \quad \text{for every } x=(x_1,x_2,x_3)\varepsilon C.$$

Let $x^0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Then $v^0 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $x^1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $v^1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $x^2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $v^2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$,

and $x^3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = x^1$.

Hence, the iterates cycle about the 3 points x^0, x^1 and x^2. Figure 4.7 illustrates this cyclic behavior.

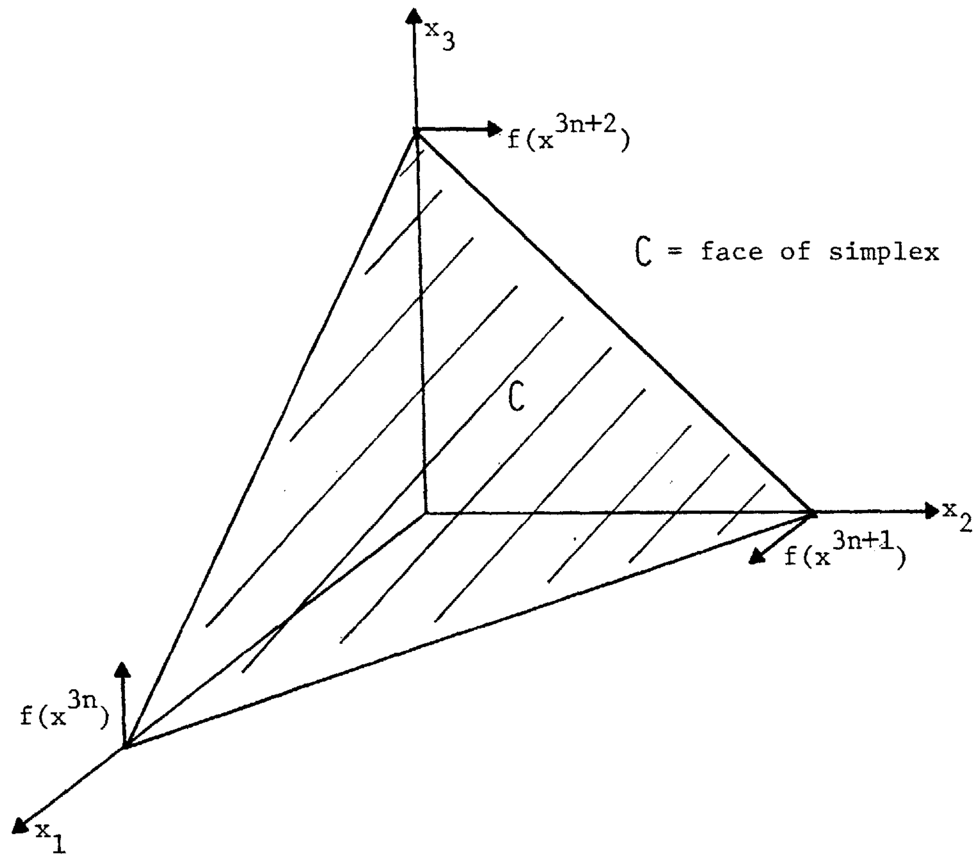


Figure 4.7 The Generalized Frank-Wolfe Method Cycles on a Traffic Equilibrium Problem

The generalized Frank-Wolfe method does not converge in the above examples because the matrix M is, in some sense, "too asymmetric". The algorithm seems to cycle, however, only when the Jacobian of f is very asymmetric. Because the generalized Frank-Wolfe algorithm reduces to the generalized steepest descent problem when the problem to which it is being applied is unconstrained, it is likely that the conditions required for the generalized Frank-Wolfe to converge are at least as strong as the condition (see Theorem 3.3) required for the generalized steepest descent method to converge. That is, it is likely that at least, M^2 must be positive definite. This condition is satisifed in neither of the previous examples. In Example 4.3, $M^2 = \begin{bmatrix} -2 & 2\sqrt{3} \\ -2\sqrt{3} & -2 \end{bmatrix}$ is clearly not positive definite.

In Example 4.4,

$$M^2 = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix} \text{ and } 2\widehat{M^2} = \begin{bmatrix} 2 & 3 & 3 \\ 3 & 2 & 3 \\ 3 & 3 & 2 \end{bmatrix} . \quad \widehat{M^2}$$

is not positive definite because $\det \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} = -5 < 0.$

Several difficulties arise when trying to prove convergence of the generalized Frank-Wolfe method. First, the iterates generated by the algorithm do not contract toward the constrained solution with respect to either the Euclidean norm or the $\hat{M}$ norm, where $M = \nabla f(x^*)$, even if M is symmetric. The following example illustrates this problem.

<u>Example 4.5</u>

Let $f(x) = x$ and let $C = \{x = (y,z): y \leq 1,\ z \leq 1 \text{ and } y + z \geq 1\}$. The solution to VI(f,C) is $x^* = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$. Let $x^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then $f(x^0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and the linear program $\min\{x^T f(x^0): x \in C\}$ has a vertex solution at $v^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The variational inequality subproblem also solves at the point v^0, so $x^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Thus, $\| x^1 - x^* \|_2 = \frac{\sqrt{2}}{2} = \| x^0 - x^* \|_2$. The iterates cannot contract to the solution in Euclidean norm (or, equivalently, in $\hat{M} = I$ norm) because $\| x^1 - x^* \|_2$ is not strictly less than $\| x^0 - x^* \|_2$. Figure 4.8 illustrates this example.

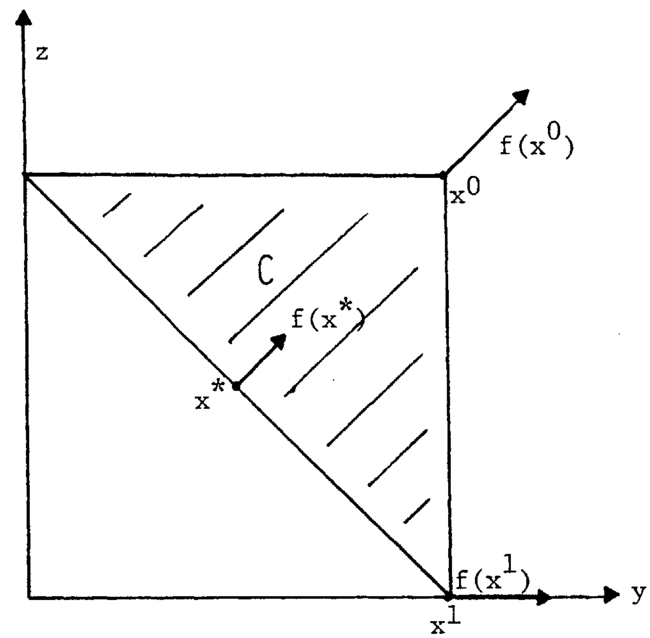
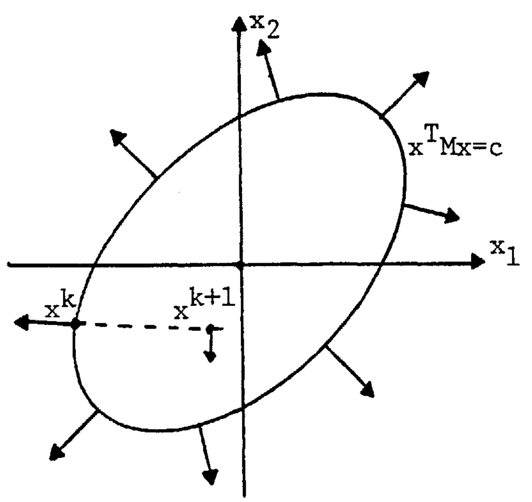


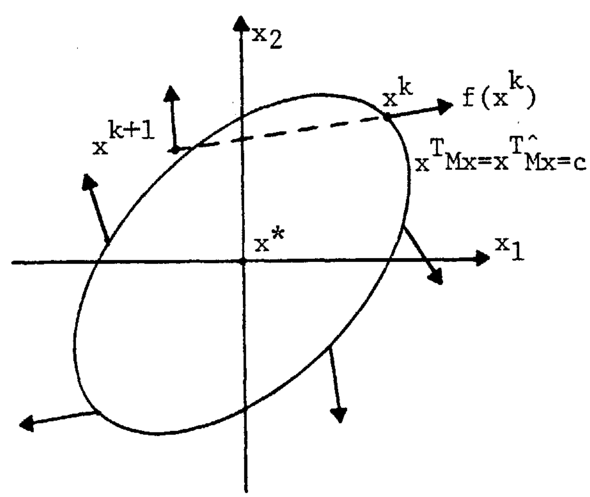
Figure 4.8 Frank-Wolfe Iterates Need Not Contract to the Solution

The proof of convergence of the Frank-Wolfe method for convex minimization problems demonstrates convergence by showing that $F(x^k)$ is a descent function. When $f(x) = Mx - b$ is a gradient mapping, instead of using the usual descent argument, we can prove convergence of the generalized Frank-Wolfe method with an argument that relies on the fact that the solution of the constrained problem is the projection onto the feasible region of the unconstrained solution with respect to the $\hat{M}$ norm. This result is not true if M is asymmetric, so the argument cannot be generalized to the asymmetric case.

When M is a symmetric positive definite matrix, the equation $x^T Mx \leq 1$ describes an ellipsoid whose axes are in the direction of the eigenvectors of M. This is true because the rotation $y=U^T x$ produces the sum of squares $x^T Mx = x^T UWU^T x = y^T Wy = \lambda_1 (y_1)^2 + \ldots + \lambda_n (y_n)^2$, where W is a diagonal matrix with diagonal entries λ_i, and $\{\lambda_i\}_{i=1}^n$ is the set of eigenvalues of $\hat{M}$. The equations $x^T Mx=c$, therefore, describe concentric ellipsoids about the origin. At each point x on the boundary of an ellipsoidal level set, the vector $f(x) = Mx$ is normal to the hyperplane supporting the set at the point x. In this case, the Frank-Wolfe iterates will get closer to the solution with respect to the M norm, because the solution x^{k+1} to the k^{th} variational inequality subproblem cannot lie on the boundary of the same level set as the previous iterate x^k. In contrast, if M is not symmetric, then the vector $f(x) = Mx$ is not normal to the hyperplane supporting the ellipsoid $\{x: x^T Mx = x^T \hat{M} x = c\}$ at x. In this case, the solution to the variational subproblem could lie outside of the level set that contains x^k on its boundary. Figure 4.9 illustrates the vector fields and ellipsoidal level sets for a symmetric matrix and an asymmetric matrix.



M Symmetric



M Asymmetric

Figure 4.9 f(x) = Mx is Normal to the Tangent Plane to the Level Set if and Only if M is Symmetric

Although the Frank-Wolfe algorithm itself does not converge for either of examples 4.3 or 4.4, the method will converge for these problems if the step length is reduced on each iteration. In particular, if we let the step length on the k^{th} iteration equal 1/k; i.e., the algorithm generates iterates by the recursion

$$x^{k+1} = x^k + \frac{1}{k+1}(v^k - x^k)$$

$$= \frac{1}{k+1} v^k + \frac{k}{k+1} x^k$$

$$= \frac{1}{k+1} \sum_{i=1}^{k} v^i,$$

then the procedure converges. Note that we can interpret this procedure as the Frank-Wolfe algorithm with stepsize 1/k or as an extreme-point averaging scheme: x^k is the average of the extreme points generated by the linear programming subproblems on the first k iterations. This variant of the Frank-Wolfe method is called the "fictitious play" algorithm when applied to zero-sum two-person games. The next subsection discusses this algorithm

4.3.1 Fictitious Play Algorithm

Robinson [1951] shows that an equilibrium solution (x^*, y^*) to a finite, two-person zero-sum game can be found using the iterative method of fictitious play. The game can be represented by its pay-off matrix $A=(a_{ij})$. Each play consists of a row player x choosing one of the m rows of the matrix while a column player y chooses one of the n columns. If the i^{th} row and the j^{th} column are chosen, the column player pays the row player the amount a_{ij}, i.e. the column player receives $-a_{ij}$ and the row player receives $+a_{ij}$.

In this subsection, we show that the Frank-Wolfe algorithm with stepsize 1/k is a generalization of the fictitious play algorithm, and establish convergence of the algorithm for a class of variational inequality problems. We start by reformulating the matrix game as a variational inequality problem. To formally define a solution to the matrix game, let S^k be the unit simplex in R^k. Then an equilibrium solution (x^*, y^*) to the game is a pair of points $x^* \varepsilon S^m$, $y \varepsilon S^n$ satisfying

$$x^T Ay^* \leq (x^*)^T Ay^* \leq (x^*)^T Ay \qquad \text{for every } x \varepsilon S^m, \ y \varepsilon S^n.$$

In the fictitious play strategy, at each iteration (i.e., for each play), each player plays the best <u>pure</u> strategy (i.e., selects a <u>single</u> best row or column) against the accumulated strategies of the other player. Hence, at iteration k, the row player x chooses the pure strategy $\bar{x}^k$ that is the best reply to the accumulated strategies played by the column player; i.e., to the average $y^k = \frac{1}{k} \sum_{j=0}^{k-1} \bar{y}^j$ of the first k plays by the column player. If $\bar{x}^k$ is the best response to y^k, $\bar{x}^k$ must satisfy

$$x^T Ay^k \leq (\bar{x}^k)^T Ay^k \qquad \text{for every } x \varepsilon S^m.$$

That is, $\bar{x}^k$ solves the (trivial) linear program

$$\underset{x \varepsilon S^m}{\text{Max}} \ x^T Ay^k.$$

Similarly, $\bar{y}^k$ solves the linear program

$$\min_{y \varepsilon S^m} \ (x^k)^T Ay, \qquad \text{where } x^k = \frac{1}{k} \sum_{i=0}^{k-1} \bar{x}^i .$$

Robinson shows that the iterates generated by this strategy converge to the equilibrium solution of the game, i.e.

$$\lim_{k \to \infty} x^k = \lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} \bar{x}_i = x^* ,$$

and

$$\lim_{k \to \infty} y^k = \lim_{k \to \infty} \frac{1}{k} \sum_{j=0}^{k-1} \bar{y}_j = y^* .$$

The saddlepoint problem, find $x^* \varepsilon S^m$, $y^* \varepsilon S^n$ satisfying

$$x^T Ay^* \leq (x^*)^T Ay^* \leq (x^*)^T Ay \qquad \text{for every } x \varepsilon S^m \text{ and } y \varepsilon S^n,$$

or, equivalently, such that

$$x^T Ay^* \leq (x^*)^T Ay \qquad \text{for every } x \varepsilon S^m \text{ and } y \varepsilon S^n,$$

can be reformulated as the following variational inequality problem VI(f,C):

$$\text{Find } z^* \varepsilon C = S^m x S^n \text{ satisfying } (z-z^*)^T f(z^*) \geq 0$$
$$\text{for every } z \, \varepsilon \, C,$$

where $z = \begin{bmatrix} x \\ y \end{bmatrix}$ and $f(z) = Mz = \begin{bmatrix} 0 & -A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -Ay \\ A^T x \end{bmatrix}$.

Then $(z-z^*)^T f(z^*) \geq 0$ for every $z \varepsilon C$ if and only if $(x-x^*)^T(-Ay^*)+(y-y^*)^T A^T x^*$ ≥ 0 for every $x \varepsilon S^m$, $y \varepsilon S^n$, which is true if and only if $(x)^T Ay^* \leq (x^*)^T Ay$ for every $x \varepsilon S^m$ and $y \varepsilon S^n$.

The fictitious play method first finds $\bar{z}^k = \begin{bmatrix} \bar{x}^k \\ \bar{y}^k \end{bmatrix}$ by solving the linear program

$$\underset{x \varepsilon C}{\text{Min}} \; z^T f(z^k).$$

(This subproblem determines $\bar{x}^k$ and $\bar{y}^k$ because

$$z^T f(z^k) = (x^T, y^T) \begin{bmatrix} -Ay^k \\ A^T x^k \end{bmatrix} = -x^T Ay^k + (x^k)^T Ay, \text{ and, hence,}$$

$\bar{z}^k$ minimizes $z^T f(z^k)$ over C if and only if $\bar{x}^k$ maximizes $x^T Ay^k$ over S^m and $\bar{y}^k$ minimizes $(x^k)^T Ay$ over S^n.) The algorithm then determines the next iterate $z^{k+1} = \frac{1}{k+1} \sum_{i=0}^{k} \bar{z}^i$. Viewing the fictitious play algorithm applied to the matrix game reformulated as a variational inequality problem suggests the following algorithm to solve the general variational inequality problem VI(f,C) when C is polyhedral.

<u>Generalized Fictitious Play Algorithm</u>

Step 0: Select $x^0 \varepsilon C$. Set $k = 0$.

Step 1: Given x^k, let y^k be a vertex solution to the linear program

$$\underset{x \varepsilon C}{\text{Min}} \; x^T f(x^k).$$

If $(y^k)^T f(x^k) = (x^k)^T f(x^k)$, stop: x^k is a solution to VI(f,C).

Otherwise, go to Step 1 with $x^{k+1} = \frac{1}{k+1} \sum_{i=0}^{k} y^i =$

$$\frac{1}{k+1} y^k + \frac{k}{k+1} x^k = x^k + \frac{1}{k+1} (y^k - x^k). \qquad \blacksquare$$

The fictitious play algorithm may be reinterpreted as the Frank-Wolfe method with averaging. At the k^{th} iteration, the Frank-Wolfe method solves the same linear program subproblem as in Step 1 above, and then moves to the point on the line segment $[x^k, y^k]$ that solves the one-dimensional variational inequality problem on the line segment, while the fictitious play algorithm moves to the point that is the average of all of the subproblem solutions $y^0,\ldots,y^k$ thus far generated. Note that both procedures move in the direction $y^k - x^k$ from the point x^k.

The following theorem shows that the fictitious play algorithm will solve a certain class of variational inequality problems. Shapley [1964] has devised an example that shows that the method of fictitious play need not solve general bimatrix games (and, hence, general variational inequality problems). However, the mapping in the example is not monotone.

Theorem 4.6

The fictitious play algorithm will produce a sequence of iterates that converge to the solution of the variational inequality problem VI(f,C) if

(i) f is continuously differentiable and monotone;

(ii) C is compact and strongly convex; and

(iii) no point x in the ground set C satisfies $f(x) = 0$.

Proof

The algorithm fits into the framework of Auslender's [1976] descent algorithm procedure (see section 2.4.4), because y^k satisfies the subproblem $\min\{x^T f(x^k): x \in C\}$ and the stepsize $w_k = 1/k$ at the k^{th} iteration satisfies $w_k > 0$, $\sum_{k\to\infty}^{\infty} w_k = +\infty$, and $\lim_{k\to\infty} w_k = 0$. Therefore, the algorithm converges if conditions (i), (ii) and (iii) are satisfied. ∎

Two of the conditions specified by the theorem are more restrictive than we might wish. First, if C is strongly convex, then C cannot be polyhedral. This framework, therefore, does not show that the algorithm converges for the many problem settings that cast the variational inequality problem over a polyhedral ground set. Since an important feature of this algorithm is that the subproblem is a linear program when the ground set C is polyhedral, this restriction renders the algorithm much less attractive. Secondly, the condition that $f(x) \neq 0$ for $x \in C$ may be too restrictive in some problem settings. One setting for which this condition is not too restrictive is the traffic equilibrium problem. If we assume that the demand between at least one OD pair is positive, then we can assume that the cost of any feasible flow on the network is nonzero.

Powell and Sheffi [1982] show that iterative methods with "fixed step sizes" such as this one will solve convex minimization problems under certain conditions. Their proof does not extend to variational inequality problems that are defined by monotone maps that have asymmetric Jacobians. Although we do not currently have a convergence proof for the fictitious play algorithm for solving variational inequality problems, we believe that it is likely that the algorithm will converge. We therefore end this section with the following conjecture:

Conjecture

If f is uniformly monotone and C is a bounded polyhedron, then the fictitious play algorithm will solve the variational inequality problem VI(f,C).

4.4 Concluding Remarks

In this chapter, we have analyzed several algorithms for solving variational inequality problems defined by monotone mappings. All of these algorithms reduce to first-order approximation algorithms when they are used to solve variational inequality problems that are equivalent to convex minimization problems.

Section 4.1 analyzes an algorithm that solves a variational inequality problem by solving a sequence of quadratic programming subproblems, or, equivalently, a sequence of affine variational inequality subproblems defined by symmetric matrices. This "contracting ellipsoid" algorithm solves a variational inequality problem if the Jacobian of the underlying mapping satisfies a condition that restricts its degree of asymmetry. The algorithm solves an unconstrained affine variational inequality problem by generating sequence of ellipsoids that contract to the solution of the problem. We show that the algorithm is closely related to matrix splitting algorithms, projection algorithms, and a general iterative algorithm devised by Dafermos.

Section 4.2 discusses a subgradient algorithm for solving a max-min problem that is equivalent to the variational inequality problem. The algorithm requires only that the mapping f be continuous and monotone and that the set C be closed and convex. We show that the subgradient algorithm solves the same subproblem as the "contracting ellipsoid" method, but that the two procedures choose different movement directions.

Section 4.3 analyzes the behavior of the Frank-Wolfe method for convex minimization problems when it is generalized to solve monotone variational inequality problems. The generalized Frank-Wolfe method need not converge when applied to a variational inequality problem if the underlying mapping is not a gradient mapping. We show, however, a variant of the Frank-Wolfe procedure, the "fictitious play algorithm," solves a certain class of variational inequality problems.

CHAPTER 5

CONCLUSION

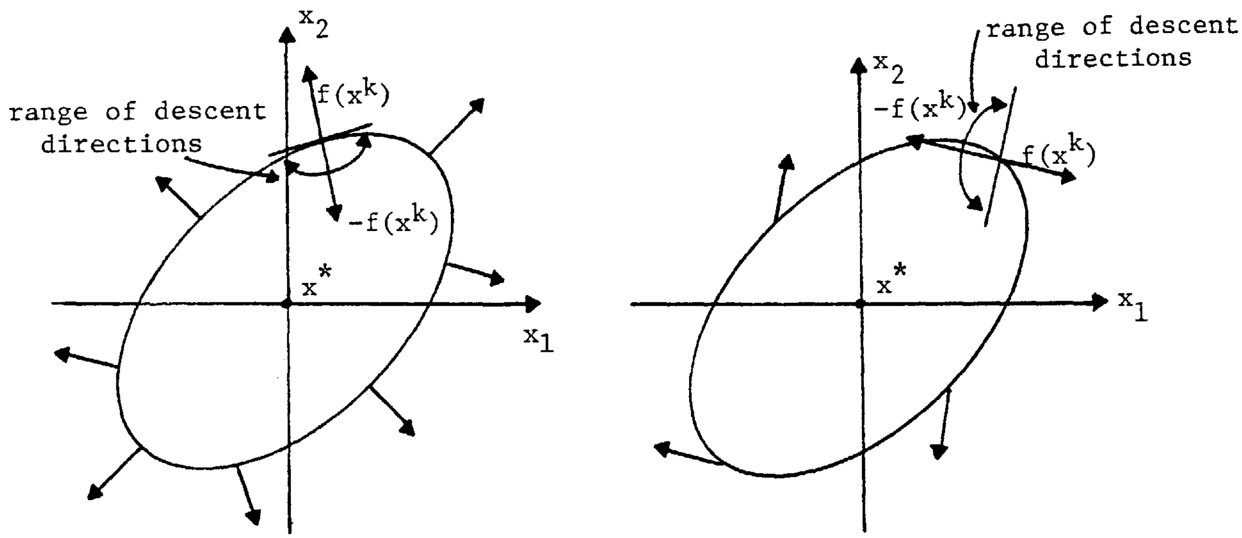
A variational inequality problem defined by a monotone mapping is a generalization of a convex minimization problem. This thesis analyzes a number of algorithms to solve variational inequality problems. The main thrust of the work is to determine when nonlinear programming algorithms can be generalized to solve variational inequality problems. A variational inequality problem is equivalent to a convex minimization problem exactly when the Jacobian of the mapping that defines the variational inequality is symmetric over the feasible set. Therefore, the type of condition that allows a nonlinear programming algorithm to solve a variational inequality problem tends to restrict the degree of asymmetry of the underlying mapping.

One such condition, that M^2, the square of the Jacobian matrix of the mapping, be positive definite, affects the convergence of several algorithms. The steepest descent algorithm, generalized to solve unconstrained variational inequality problems defined by strictly monotone affine maps, converges if and only if M^2 is positive definite. In addition, the contracting ellipsoid algorithm solves unconstrained variational inequality problems if M^2 is positive definite, and, it is likely that M^2 positive definite is a necessary condition for the generalized Frank-Wolfe method to converge.

The reason that the convergence of nonlinear programming algorithms adapted to solve variational inequality problems requires a restriction on the degree of asymmetry of the Jacobian is that, in general, nonlinear programming algorithms iteratively move in "good" feasible descent directions. That is, for the minimization problem

$$\min_{x \in C} F(x),$$

for $k = 0,1,\ldots$, the algorithms determine a feasible direction d_k satisfying $d_k^T \nabla F(x^k) < 0$. Many algorithms attempt to choose d_k "as close as possible" to the steepest descent direction $-\nabla F(x^k)$. When these algorithms are adapted to solve variational inequality problems, they determine on the k^{th} iteration a direction d_k satisfying $d_k^T f(x^k) < 0$, with d_k as close as possible to the steepest "descent" direction, $-f(x^k)$. As long as the Jacobian of f is nearly symmetric, such a direction is a "good" direction for the problem VI(f,C), because a move in the direction d_k is a move towards the solutions. If, however, the Jacobian of f is very asymmetric, a move in the direction d_k may be a move away from the solution. Figure 5.1 illustrates the set of "descent" directions for two problems, one defined by an affine map with a positive definite symmetric matrix and one defined by an affine map with a positive definite asymmetric matrix. The illustrations show that $-f(x^k)$, the direction that a nonlinear programming algorithm considers to be the "best" direction, can be a poor direction if the matrix is very asymmetric.



M Symmetric

M Asymmetric

Figure 5.1 The Direction $-f(x^k)$ is not a Good Movement Direction if the Jacobian of f(x) is Very Asymmetric

Projection algorithms are one of the most widely used procedures to solve variational inequality problems. A fundamental difference between most of the nonlinear programming algorithms that we consider in this work and projection methods is that we consider algorithms that use a "full steplength"; in contrast, projection methods use a small fixed steplength, or a steplength defined by a convergent sequence of real numbers. Although these full steplength algorithms, such as the generalized steepest descent and Frank-Wolfe algorithms and the contracting ellipsoid method, require more work per iteration that those using a constant or convergent sequence step size, they move fairly quickly to a neighborhood of the solution. Taking a full steplength poses a problem, however, when the Jacobian of the mapping is very asymmetric. In this case, the "twisting" vector field may not only cause the algorithm to choose a less than ideal direction of movement, but, having done so, will cause the algorithm to determine a much longer step than it would choose if the mapping was nearly symmetric. This asymmetry is not as much of a problem is the step size is small, because the algorithm will not pull as far away from the solution even if the direction of movement is poor. Figure 5.2 illustrates the affect of asymmetry on the full steplength.

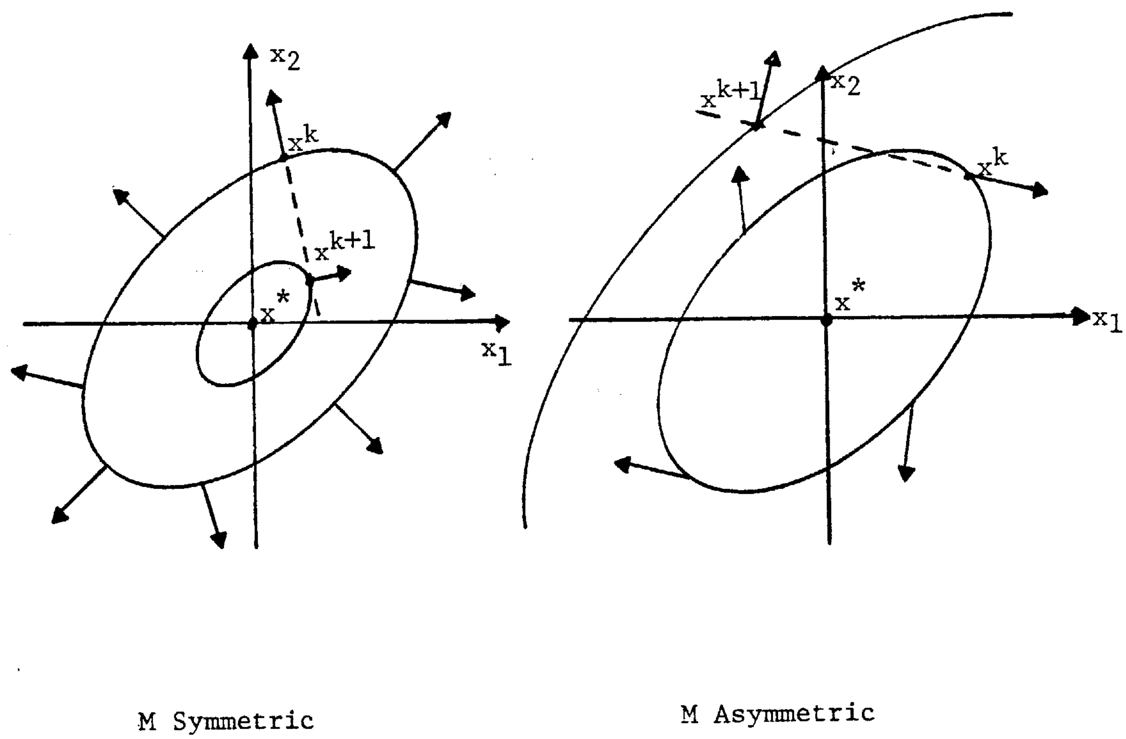


Figure 5.2

A Full Steplength Pulls the Iterate Further From the Solution When the Map is Very Asymmetric

By our previous observations, algorithms that take a full step size will converge only if a bound on the degree of asymmetry of the Jacobian is imposed. Projection methods do not require this type of condition. Most of the algorithms that we consider in this work will converge even if the monotone problem mapping is very asymmetric as long as the full steplength is replaced by a sufficiently small steplength. The steepest descent algorithm for unconstrained variational inequality problems becomes a projection algorithm if the stepsize is sufficiently small. Theorem 4.6 shows that the Frank-Wolfe method will converge for a class of variational inequality problems if the stepsize is defined by a convergent sequence. Recall also, Lemma 4.2, that the contracting ellipsoid algorithm for variational inequality problems defined by affine maps can be considered as a projection algorithm with a step size of one at each iteration. Thus, it also will converge even if the conditions on the symmetry of the Jacobian are not satisfied as long as the step size is sufficiently small.

We see a number of directions for future research. The convergence of the generalized Frank-Wolfe method and the fictitious play algorithm for problems defined over polyhedral sets remain important open problems. Secondly, we would like to further investigate procedures for scaling problem maps so that convergence conditions are satisfied and to accelerate convergence. We would also like to further investigate generalized descent algorithms. It may be possible to avoid some of the previously mentioned problems with asymmetric maps by using these methods to determine better movement directions. Another interesting research problem is to analyze how much accuracy is lost by solving a problem that approximates a mapping that has a nearly symmetric Jacobian by a mapping that has a symmetric Jacobian.

Conceivably, the major contribution of this work is conceptual: it enhances our understanding of a certain class of problems. Although the contracting ellipsoid method should, because of its more complex subproblem, be able to efficiently solve a wide class of problems, the steepest descent and Frank-Wolfe algorithms are likely to be good algorithms only for problems that have nearly symmetric Jacobians. This work mainly contributes to the understanding of the role of symmetry of the Jacobian in the convergence of nonlinear programming algorithms generalized to solve variational inequality problems.

REFERENCES

Aashtiani, H.Z. [1979]. "The Multi-Modal Traffic Assignment Problem," Ph.D. Thesis, Sloan School of Management, M.I.T., Cambridge, MA.

Aashtiani, H.Z. and T.L. Magnanti [1980]. "Equilibria on a Congested Transportation Network," SIAM Journal on Algebric and Discrete Methods 2:3, 213-226.

Aashtiani, H.Z. and T.L. Magnanti [1982]. "A Linearization and Decomposition Algorithm for Computing Urban Traffic Equilibria," Proceedings of the 1982 IEEE Large Scale Systems Symposium, Oct. 1982.

Ahn, B. [1979]. "Computation of Market Equilibria for Policy Analysis: The Project Independence Evaluation System Approach," Ph.D. dissertation, Department of Engineering - Economic Systems, Stanford University, Stanford, California, and Garland Publishing, Inc. New York, NY.

Ahn, B. and W. Hogan [1982]. "On Convergence of the PIES Algorithm for Computing Equilibria," Operations Research 30:2, 281-300.

Anstreicher, K. [1984]. Private communication.

Auslender, A. [1976]. Optimisation: Méthodes Numbériques, Mason, Paris.

Avriel, M. [1976]. Nonlinear Programming: Analysis and Methods, Prentice-Hall, Englewood Cliffs, N.J.

Bakušinskiĭ, A.B. [1979]. "Equivalent Transformations of Variational Inequalities and their Use," Soviet Mathematics Doklady 20:4, 877-880.

Bakušinskiĭ, A.B. and B.T. Poljak [1974]. "On the Solution of Variational Inequalities," Soviet Mathematics Doklady 15, 1705-1710.

Beckmann, M.J., C.B. McGuire and C.B. Winsten [1956]. Studies in the Economics of Transportation, Yale University Press, New Haven, CT.

Bertsekas, D.P. [1982]. Constrained Optimization and Lagrange Multiplier Methods, Academic Press, New York, NY.

Bertsekas, D.P. and E.M. Gafni [1980]. "Projection Methods for Variational Ineqalities with Application to the Traffic Assignment Problem," LIDS, M.I.T., Cambridge, MA.

Browder, F.E. [1966]. "Existence and Approximation of Solutions of Nonlinear Variational Inequalities," Proceedings of the National Academy of Sciences, 56, 1080-1086.

Bruynooghe, M., A. Gibert, and M. Sakarovitch [1968]. "Une Méthode d'Affectation du Traffic," in Fourth Symposium of the Theory of Traffic Flow, Karlsruhe.

Cantor, D. and Gerla, M. [1974]. "Optimal Routing in a Packet Switched Computer Network," IEEE Transactions on Computers 10, 1062-1068.

Cauchy, A.L. [1847]. "Méthode générale pour la résolution des systèmes d'équations simultanées," Comptes rendus, Ac. Sci. Paris 25, 536-538.

Cottle, R.W. [1974]. "Complementarity and Variational Problems," Symposia Mathematica XIX, Academic Press, New York, NY, 59-70.

Cottle, R.W., G.H. Golub and R.S. Sacher [1978]. "On the Solution of Large Structured Linear Complementarity Problems: The Block Partitioned Case," Applied Mathematics and Optimization 4, 347-363.

Cottle, R.W., F. Giannessi, and J-L. Lions, eds. [1980]. Variational Inequalities and Complementarity Problems. John Wiley and Sons, New York, NY.

Courant, R. [1943]. "Variational Methods for the Solution of Problems of Equilibrium and Vibrations," Bull. Amer. Math. Soc. 49, 1-23.

Curry, H. [1944]. "The Method of Steepest Descent for Nonlinear Minimization Problems," Quar. Appl. Math. 2, 258-261.

Dafermos, S.C. [1971]. "An Extended Traffic Assignment Model with Applications to Two-Way Traffic," Transportation Science 4, 336-389.

Dafermos, S.C. [1972]. "The Traffic Assignment Problem for Multiclass-User Transportation Networks," Transportation Science 6, 73-87.

Dafermos, S.C. [1980]. "Traffic Equilibrium and Variational Inequalities," Transportation Science 14, 43-54.

Dafermos, S.C. [1982a]. "Relaxation Algorithms for the General Asymmetric Traffic Equilibrium Problem," Transportation Science 16:2, 231-240.

Dafermos, S.C. [1982b]. "The General Multimodal Network Equilibrium Problem with Elastic Demands," Networks 12:1, 57-72.

Dafermos, S.C. [1983]. "An Iterative Scheme for Variational Inequalities," Mathematical Programming, 26:1, 40-47.

Eaves, C. [1978a]. "A Locally Quadratically Convergent Algorithm for Computing Stationary Points," Technical Report, Department of Operations Research, Stanford University, Stanford, California.

Eaves, C. [1978b]. "Computing Stationary Points," Mathematical Programming Study 7, 1-14.

Eaves, C. [1978c]. "Computing Stationary Points, Again," in O.L. Mangasarian, R.R. Meyer and S.M. Robinson eds., Nonlinear Programming 3, Academic Press, NY, 391-405.

Fisk, C.S. and S. Nguyen [1982]. "Solution Algorithms for Network Equilibrium Models with Asymmetric User Costs," Transportation Science 16:3, 361-381.

Fisk, C. and D.E. Boyce [1982]. "A General Variational Inequality Formulation of the Network Equilibrium-Travel Choice Problem," Publication No. 7, Transportation Planning Group, Department of Civil Engineering, University of Illinois at Urbana-Champaign.

Fisk, C.S. and D.E. Boyce [1983]. "Alternative Variational Inequality Formulations of the Network Equilibrium-Travel Choice Problem," Transportation Science 17:4, 454-463.

Florian, M.A. (ed.) [1976]. Traffic Equilibrium Methods, Lecture Notes in Economics and Mathematical Systems 118, Springer-Verlag, New York, NY.

Florian, M. and M. Los [1981]. "A New Look at Static Spatial Price Equilibrium Models," Centre de Recherche sur les Transports, Publication #196, Université de Montréal, Montréal, Québec.

Florian, M. and H. Spiess [1982]. "The Convergence of Diagonalization Algorithms for Asymmetric Network Equilibrium Problems," Transportation Research B 16B, 477-483.

Frank, M. and P. Wolfe [1956]. "An Algorithm for Quadratic Programming," Naval Research Logistic Quarterly 3, 95-110.

Geoffrion, A.M. [1971]. "Elements of Large-scale Mathematical Programming," Management Science 16, 652-691.

Gershgorin, S. [1931], "Über die Abrenzung der Eigenwerte einer Matrix," Izv. Akad. Nauk SSSR Ser. Mat. 7, 749-754.

Golden, B. [1975]. "A Minimum Cost Multi-Commodity Network Flow Problem Concerning Imports and Exports. Networks 5, 331-356.

Hadamard, J. [1908]. "Mémoire sur le problème d'analyse relatif à l'équilibre des plaques élastiques encastrées," Mémoires presentés par divers savants éstranges à l'Academie des Sciences de l'Institut de France(2) 33:4.

Hartman, P. and G. Stampacchia [1966]. "On Some Nonlinear Elliptic Differential Functional Equations," Acta Mathematica 115, 271-310.

Hearn, D. [1982]. "The Gap Function of a Convex Program," Operations Research Letters 1:2, 67-71.

Hearn, D., S. Lawphongpanich and S. Nguyen [1983]. "Convex Programming Formulations of the Asymmetric Traffic Assignment Problem," Research Report No. 83-12, Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida.

Hearn, D. and S. Nguyen [1982]. "Dual and Saddle Functions Related to the Gap Function," Research Report No. 82-4, Deparment of Industrial and Systems Engineering, University of Florida, Gainesville, Florida.

Holloway, C.A. [1974]. "An Extension of the Frank-Wolfe Method of Feasible Directions," Mathematical Programming 6:1, 14-27.

Karamardian, S. [1971]. "Generalized Complementarity Problem," Journal of Optimization Theory and Applications 8, 161-168.

Khachiyan, L.G. [1979]. "A Polynomial Algorithm in Linear Programming," Soviet Mathematics Doklady, 20:1, 191-194.

Kinderlehrer, D. and Stampacchia [1980]. An Introduction to Variational Inequalities and Applications, Academic Press, New York, NY.

Korpelevich, G.M. [1977]. "The Extragradient Method for Finding Saddle Points and Other Problems," Matekon 13:4, 35-49.

Lawphongpanich, S. and D. Hearn [1982]. "Simplicial Decomposition of the Asymmetric Traffic Assignment Problem," Research Report No. 82-12, Department of Industrial and Systems Engineering, University of Florida, Gainsville, Florida.

LeBlanc, L.J., E.K. Morlok, and W.P. Pierskalla [1975]. "An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem," Transportation Research 5, 309-318.

Lemke, C.E. [1980]. "A Survey of Complementarity Theory," Chapter 15 in Cottle et al. [1980], 213-239.

Leventhal, T., G. Nemhauser, and L. Trotter [1973]. "A Column Generation Algorithm for Optimal Traffic Assignment," Transportation Science 7:2, 168-172.

Luenberger, D.G. [1973]. Introduction to Linear and Nonlinear Programming. Addison-Wesley, Reading, MA.

Luque, F.J. [1984]. "Asymptotic Convergence Analysis of the Proximal Point Algorithm," SIAM Journal on Control and Optimization 22:2, 277-293.

Luthi, H-J. [1983]. "On the Solution of Variational Inequalities by the Ellipsoid Method," Technical Report, Institute for Operations Research, Federal Institute of Technology, Zurich, Switzerland.

Magnanti, T.L. [1982]. "Models and Algorithms for Predicting Urban Traffic Equilibrium," Transportation Planning Models, edited by M. Florian, North-Holland (to appear).

Marcotte, Patrice [1983]. "A New Algorithm for Solving Variational Inequalities Over Polyhedra, with Application to the Traffic Assignment Problem," Cahier du GERAD #8324, École des Hautes Études Commerciales, Université de Montréal, Montréal, Québec.

Martos, B. [1975]. Nonlinear Programming Theory and Methods. American Elsevier Publishing Co., Inc., New York, New York.

Moré, J.J. [1974]. "Coercivity Conditions in Nonlinear Complementarity Problems," SIAM Review 16:1, 1-16.

Nguyen, S. [1974]. "An Algorithm for the Traffic Assignment Problem," Transportation Science 8, 203-216.

Nguyen, S. and C. Dupuis [1981]. "Une Méthode Efficace de Calcul d'un Trafic D'Équilibre dans le Cas des Coûts Non-symétriques, Centre de Recherche sur les Transports, Publication #205, Université de Montréal, Montréal, Québec.

Ortega, J.M. and W.C. Rheinboldt [1970]. Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, NY.

Pang, J.-S. [1981]. "A Column Generation Technique for the Computation of Stationary Points," Mathematics of Operations Research 6:2, 213-224.

Pang, J.-S. [1981]. "An Equivalence Between Two Algorithms for Quadratic Programming," Mathematical Programming 20, 152-165.

Pang, J.-S. and D. Chan [1981]. "Iterative Methods for Variational and Complementarity Problems," Mathematical Programming 24, 284-313.

Pang, J.-S. and D. Chan [1982]. "Gauss-Seidel Methods for Variational Inequaltiy Problems of Product Sets", Technical Report, GSIA, Carnegie-Mellon University.

Pang J.-S. and C.-S. Yu [1982]. "Linearized Simplicial Decomposition Methods for Computing Traffic Equilibria over Networks," Technical Report, University of Texas at Dallas, Dec. 1982.

Polak, E. [1971]. Computational Methods in Optimization, Academic Press, New York, NY.

Polyak, B.T. [1967]. "A General Method for Solving Extremal Problems," Dokl. Akad. Nauk SSSR 174:1, 33-36.

Polyak, B.T. [1969]. "Minimization of Unsmooth Functionals," U.S.S.R. Computational Mathematics and Mathematical Physics 9, 14-29.

Powell, W.B. and Y. Sheffi [1982]. "The Convergence of Equilibrium Algorithms with Predetermined Step Sizes," Transportation Science 16:1, 45-55.

Robinson, J. [1951]. "An Iterative Method of Solving a Game," Annals of Mathematics 54:2, 296-301.

Rockafeller, R.T. [1980]. "Lagrange Multipliers and Variational Inequalities," Chaper 20 in Cottle et. al. [1980], 303-322.

Samulson, P.A. [1952]. "Spatial Price Equilibrium and Linear Programming," American Economic Review 42, 283-303.

Shapley, L.S. [1964]. "Some Topics in Two Person Games," Advances in Game Theory, Annals of Mathematics Studies No. 52, M. Dresher, L.S. Shapley, A.W. Tucker eds., Princeton University Press, Princeton, NJ, 1-28.

Shor, N.Z. [1964]. "On the Structure of Algorithms for the Numerical Solution of Optimal Planning and Design Problems," Diss. kand. fiz.-matem. n., Kiev, In-t kibernetiki AN USSR.

Sibony, M. [1970]. "Méthodes Itératives pour les Équations et Inéquations aux Dérivées Partielles Nonlinéares de Type Monotone," Calcolo 7, 65-183.

Smith, M. [1979]. "The Existence, Uniqueness and Stability of Traffic Equilibria," Transportation Research B 13B, 295-304.

Smith, M. [1983a]. "The Existence and Calculation of Traffic Equilibria," Transportation Research B 17B:4, 291-303.

Smith, M. [1983b]. "An Algorithm for Sovling Asymmetric Equilibrium Problems with a Continuous Cost-Flow Function," Transportation Research B 17B:5, 365-371.

Strang, G. [1976]. Linear Algebra and Its Applications. Academic Press, New York, N.Y.

Takayama, T. and G.G. Judge [1971]. Spatial and Temporal Price and Allocation Models, North-Holland, Amsterdam.

van de Panne, C. and A. Whinston [1969]. "The Symmetric Formulation of the Simplex Method for Quadratic Programming," Econometrica 37, 507-527.

Varga, R. [1962]. Matrix Iterative Analysis. Prentice Hall, Englewood Cliffs, NJ.

Von Hohenbalken, B. [1977]. "Simplicial Decomposition in Nonlinear Programming Algorithms,: Mathematical Programming 13, 49-68.

Wardrop, J.G. [1952]. "Some Theoretical Aspects of Road Traffic Research'" Proc. Inst. Civil Engineers, Part II 1, 325-378.

Zangwill, W.J. [1969]. Nonlinear Programming: A Unified Approach. Prentice-Hall, Englewood Cliffs, NJ.

Zuhovickii, S.I., R.A. Polyak, and M.E. Primak [1969]. "Two Methods of Search for Equilibrium Points of n-Person Concave Games," Soviet Mathematics Doklady 10:2, 279-282.