OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

A Contracting Ellipsoid Method
For Variational Inequality Problems+

by

Janice H. Hammond*
Thomas L. Magnanti**

OR 160-87 April 1987

* Harvard Business School, Boston, MA 02163
** Sloan School of Management, MIT, Cambridge, MA 02139

+ This research has been supported in part by Grant #ECS-83-16224 from the Operations Research and Systems Theory Program of the National Science Foundation

A CONTRACTING ELLIPSOID METHOD FOR VARIATIONAL INEQUALITY PROBLEMS

by

Janice H. Hammond
Thomas L. Magnanti

ABSTRACT

A variational inequality defined by a symmetric map can be solved as an equivalent optimization problem. We consider an approach to solve asymmetric variational inequalities that generates a sequence of variational inequality subproblems, each of which is defined by a symmetric affine map, and hence is equivalent to a quadratic program. We interpret the algorithm geometrically in terms of a sequence of contracting ellipsoids that are implicitly generated by the algorithm, and interpret the convergence conditions as near-symmetry conditions imposed upon the underlying map. We discuss connections between this algorithm and its geometry and several related methods for solving variational inequalities.

Keywords: variational inequalities, first order methods, asymmetric maps

1. Introduction

The variational inequality problem VI(f,C) seeks a solution to a specially structured system of inequalities: namely, for a given set $C \subseteq R^n$ and mapping $f: C \to R^n$, the problem seeks a solution $x^* \in C$ that satisfies

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

This general problem formulation encompasses a wide range of problem types. It 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.

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. (See, for example, Hartmann and Stampacchia [1966], Browder [1966], Lions and Stampacchia [1967], Sibony [1970], and Kinderlehrer and Stampacchia [1980]). In these settings, the problem is usually formulated over an infinite dimensional function space; in contrast, this paper discusses problems formulated over finite dimensional spaces.

Recent results concerning network equilibrium problems have heightened mathematical programmers' interest in variational inequalities. In particular, the recognition (see Smith [1979] and Dafermos [1980] and related work by Asmuth [1978] and Aashtiani and Magnanti [1981]) that the equilibrium conditions for urban traffic equilibria can be formulated in a natural way as a variational inequality (or nonlinear complementarity) problem and the desire to find methods to solve such equilibrium problems have motivated a number of researchers to develop algorithms to solve variational inequalities. (See, for example, Ahn and Hogan [1982], Auslender [1976], Dafermos [1983], Hearn [1982], Harker [1986], Marcotte [1985], Pang and Chan [1981], and Pang [1985]. Hammond [1984] provides a survey of these and other papers on algorithms for solving variational inequality problems.)

In this paper we introduce and analyze a "contracting ellipsoid" algorithm for solving variational inequalities and discuss its relationship with two related variational inequality algorithms: a generalized steepest descent algorithm and a subgradient algorithm. The contracting ellipsoid algorithm has a simple, yet revealing, underlying geometric structure. This structure not only captures the relationship between the vector field defined by the problem map and the course the algorithm takes, but also provides a framework that aids in the understanding of related "generalized descent" algorithms. Throughout our discussion we emphasize these interpretations. Indeed, we see these geometrical insights as a major component of the paper.

2. First-Order Approximation Methods

Many algorithms for solving 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 paper we discuss several variational inequality algorithms, each of which generalizes a first-order approximation algorithm for a nonlinear programming problem: that is, when the variational inequality problem reduces to an equivalent optimization problem, each method reduces to a first-order approximation algorithm that solves that optimization problem. To state the conditions under which VI(f,C) reduces to an equivalent optimization problem, we first recall that a continuously differentiable monotone mapping f is associated with a convex map $F:C \to R$ that satisfies $f(x) = \nabla F(x)$ for every $x \in C$ if and only if the Jacobian ∇f of f is symmetric on C. When f satisfies this condition, the variational inequality system (1) can be viewed as the necessary and sufficient optimality conditions for the optimization problem

$$\min \{F(x): x \in C\}. \qquad (2)$$

That is, VI(f,C) is equivalent to the convex minimization problem (2) exactly when the continuously differentiable map f is monotone and has a symmetric Jacobian on C.

At each iteration, a first-order approximation algorithm for solving problem (2) approximates F by a function depending on the gradient of F. Linear approximation methods, which are classical examples of such methods, generate an iterate x^{k+1} from the previous iterate x^k based on $F^k(x)$, the linear approximation to F about x^k defined by

$$F^k(x) := F(x^k) + \nabla F(x^k)(x - x^k). \qquad (3)$$

For example, given x^k, the Frank-Wolfe method chooses as the next iterate a point x^{k+1} that minimizes F on the line segment from x^k to v^k, where v^k is the solution to the subproblem

$$\min \{F^k(x): x \in C\}.$$

The steepest descent algorithm can also be viewed as a first-order approximation method: given x^k, the algorithm chooses a point x^{k+1} that minimizes F in the direction d^k, where d^k is the solution to the subproblem

$$\min \{F^k(x): ||x|| \le 1\}.$$

A more accurate first-order approximation would replace the (constant) gradient vector $\nabla F(x^k)$ by the (nonlinear) gradient vector $\nabla F(x)$, giving the approximation

$$\hat{F}^k(x) := F(x^k) + \nabla F(x)(x - x^k). \qquad (4)$$

In this paper, we investigate variations of both of these first-order approximation schemes as adapted to solve variational inequality problems. In Section 3 we summarize the convergence properties of a generalized steepest descent algorithm. Sections 4 and 5 analyze variational inequality algorithms that generalize first-order approximation methods based on the approximation (4). Our analysis of a "contracting ellipsoid" algorithm in Section 4 provides a geometrical framework within which to view a number of variational inequality algorithms. In Section 5, we discuss 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.

3. A Generalized Steepest Descent Algorithm

In this section, we summarize results concerning a generalized steepest descent algorithm for asymmetric systems of equations, which we view as unconstrained variational inequality problems, and show that the algorithm's convergence requires a restriction on the degree of asymmetry of the problem map. (Hammond [1984] and Hammond and Magnanti [1985] provide proofs of the results in this section and examples that illustrate the conditions under which the algorithm converges.) In the following sections, in conjunction with our discussion of a contracting ellipsoid algorithm, we further discuss the convergence conditions from this section and provide geometrical interpretations of the steepest descent algorithm.

Consider the unconstrained variational inequality problem $VI(f,R^n)$ defined by a continuously differentiable and uniformly monotone mapping f. This unconstrained problem seeks a zero of f, since $(x - x^*)^T f(x^*) \geq 0$ for every $x \in R^n$ if and only if $f(x^*) = 0$.

The following algorithm generalizes the well-known steepest descent algorithm for convex minimization problems: when f is the gradient of F, the algorithm reduces to the usual steepest descent algorithm applied to (2). (In the statement of the algorithm, $[x;y]$ denotes the ray from x in the direction y.)

<u>Generalized Steepest Descent Algorithm for the Unconstrained Variational Inequality Problem</u>

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

Step 1: <u>Direction Choice</u>. Compute $-f(x^k)$. If $f(x^k) = 0$, stop: $x^k = x^*$. Otherwise, go to Step 2.

Step 2: <u>One-Dimensional Variational Inequality</u>. Find $x^{k+1} \in [x^k;-f(x^k)]$ satisfying $(x - x^{k+1})^T f(x^{k+1}) \geq 0$ for every $x \in [x^k;-f(x^k)]$. Go to Step 1 with $k = k + 1$. ▒

The algorithm can be viewed as a method that moves through the "vector field" defined by f by solving a sequence of one-dimensional variational inequalities. On the k^{th} iteration, the algorithm moves in the $-f(x^k)$ direction to the point x^{k+1} that satisfies $f^T(x^k)f(x^{k+1}) = 0$. If $f = \nabla F$, this orthogonality condition is equivalent to the condition that $x^{k+1} = \text{argmin}\,\{F(x): x \in [x^k;-\nabla F(x^k)]\}$, and the algorithm becomes the usual steepest descent method.

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 theorem summarizes the convergence properties of the algorithm. (In the statement of the theorem, $\hat{M} = \frac{1}{2}(M+M^T)$ denotes the symmetric part of the matrix M.)

<u>Theorem 1</u>

Let $f:R^n \to R^n$ be uniformly monotone and twice differentiable.

(a) 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 to the solution x^*, the sequence of iterates produced by the generalized steepest descent algorithm contracts in $\hat{M}$ norm to the solution.

(b) If $f(x) = Mx - b$ is an affine map (and thus, by our previous assumption, M is positive definite), then the sequence of iterates produced by the generalized steepest descent method is guaranteed to contract in $\hat{M}$ norm to the solution x^* of the problem $VI(f,R^n)$ if and only if the matrix M^2 is positive definite. Furthermore, the contraction constant is given by

$$r = \left\{ 1 - \inf_{x \neq 0} \left[\frac{[(Mx)^T(Mx)][x^T M^2 x]}{[x^T Mx][(Mx)^T M(Mx)]} \right] \right\}^{1/2} .$$

Theorem 1 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: $M^2 = M^TM$ 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 of asymmetry of M, that is, on the degree to which M can differ from M^T. To see this, note that M^2 is positive definite if and only if

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

Thus, M^2 is positive definite if and only if the angle between the vectors M^Tx and Mx is acute for every nonzero vector 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. However, if M^2 is positive definite, then the normalized quantity $||M - M^T||_2/||M + M^T||_2$ must be less than 1, where $||\cdot||_2$ denotes the Euclidean norm.

The following result establishes easily verified conditions on the matrix M that will ensure that the matrix M^2 is positive definite.

<u>Theorem 2</u>

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

$$\sum_{j\neq i} |M_{ij}| < ct \quad \text{and} \quad \sum_{j\neq i} |M_{ji}| < ct, \tag{5}$$

where $t = \dfrac{\min\{(M_{ii})^2:\ i=1,\ldots,n\}}{\max\{M_{ii}:\ i=1,\ldots,n\}}$ and $c = \sqrt{2} - 1$, then both M and M^2 are doubly diagonally dominant, and therefore positive definite, matrices.

The conditions that Theorem 2 imposes on the off-diagonal elements of M are least restrictive when the diagonal elements of M are all equal. By scaling either the rows or the columns of M in an appropriate manner before applying the generalized steepest descent algorithm, we can weaken, in some cases considerably, the convergence conditions that Theorem 2 imposes on M. Hammond and Magnanti [1985] describe the details of such scaling procedures.

4. 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. Suppose that f is the gradient of a convex function $F:R^n \rightarrow R^1$ (so that $VI(f,C)$ is equivalent to the convex minimization problem (2)), and consider an algorithm that minimizes F over C by successively minimizing the approximation $\hat{F}^k(x)$ to $F(x)$ given in (4). That is, the algorithm generates a sequence of iterates $\{x^k\}$ by the recursion

$$x^{k+1} = \text{argmin } \{\hat{F}^k(x): x \in C\} = \text{argmin } \{\nabla F(x)(x - x^k): x \in C\}. \quad (6)$$

By replacing $[\nabla F]^T$ with the mapping f in equation (6), we obtain the following algorithm that is applicable to any variational inequality problem.

<u>Contracting Ellipsoid Algorithm</u>

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

Step 1: Select $x^{k+1} \in \text{argmin } \{(x - x^k)^T f(x): x \in C\}$

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 defined by monotone affine maps, the algorithm produces a sequence of ellipsoids that contract to the solution. This algorithm is not related to Khachiyan's [1979] ellipsoid algorithm for linear programming.)

To motivate the analysis of this algorithm for the problem VI(f,C), we first consider, in Section 4.1, 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 4.2 extends these results to constrained problems defined by affine maps. Section 4.3 analyzes the algorithm for the constrained variational inequality problem defined by a nonlinear, strictly monotone mapping f. In Section 4.4, 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 4.5 discusses relationships between the contracting ellipsoid method and a number of well-known algorithms for variational inequality problems.

4.1 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 $f(x) = Mx - b$, and that M is a positive definite $n \times n$ matrix.

In this case, the minimization suoproblem

$$\min \{(x - x^k)^T f(x): x \in R^n\} \qquad (7)$$

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. These optimality conditions require that x^{k+1} satisfy

$$(M + M^T)x^{k+1} - (M^T x^k + b) = 0.$$

Hence, if $S = M + M^T$, then x^{k+1} is given by

$$x^{k+1} = S^{-1}(M^T x^k + b), \qquad (8)$$

or equivalently,

$$\begin{aligned} x^{k+1} &= x^k - S^{-1}(Mx^k - b) \\ &= x^k - S^{-1}f(x^k). \end{aligned}$$

Before proceeding with a convergence analysis, we illustrate the mechanics of the algorithm in an example.

Example 1

Let $M = \begin{bmatrix} 1 & 2 \\ -2 & 4 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Then $x^* = M^{-1}b = \begin{bmatrix} 1/4 \\ 3/8 \end{bmatrix}$, and the algorithm generates iterates by the relation (8), with

$S = M + M^T = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$. Let $x^0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then $x^1 = \begin{bmatrix} 1 \\ 3/8 \end{bmatrix}$,
$x^2 = \begin{bmatrix} 5/8 \\ 9/16 \end{bmatrix}$, $x^3 = \begin{bmatrix} 1/4 \\ 9/16 \end{bmatrix}$, $x^4 = \begin{bmatrix} 1/16 \\ 15/32 \end{bmatrix}$, $x^5 = \begin{bmatrix} 1/16 \\ 3/8 \end{bmatrix}$,
$x^6 = \begin{bmatrix} 5/32 \\ 21/64 \end{bmatrix}$, ... Figure 1 illustrates a sequence of iterates $\{x^k\}$ as well as a sequence of ellipses $\{E_0^k\}$ that we describe next.

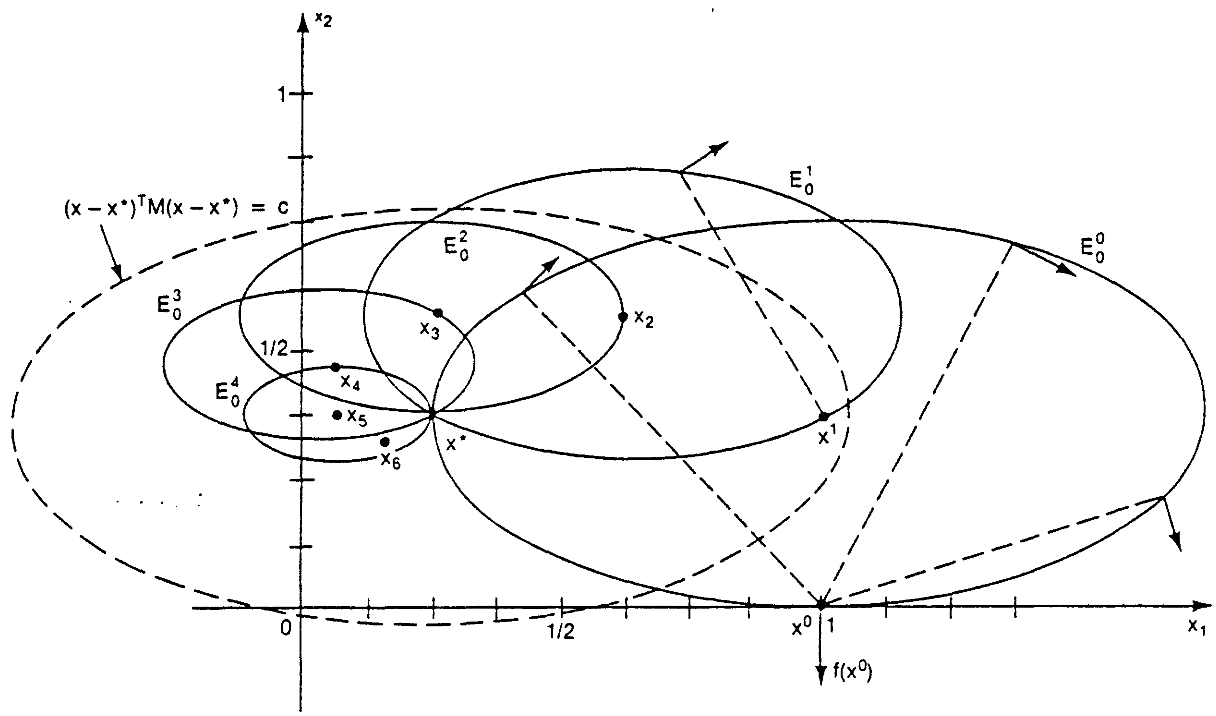


Figure 1: The Contracting Ellipsoid Method Solves An Unconstrained Affine Variational Inequality Problem by Generating Sequence of Ellipsoids that Contract to the Solution x^*.

The behavior of the algorithm can be described geometrically by considering the level sets of the objective function $(x - x^k)^T f(x)$ of the k^{th} subproblem. When f is affine, the level set

$$E_\alpha^k := \{x : (x - x^k)^T f(x) \le \alpha\}$$

is an ellipsoid centered about the point that minimizes the objective function of the k^{th} subproblem. That is, E_α^k is centered about the point $x^{k+1} = S^{-1}(M^T x^k + b)$.

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 (because $(x^k - x^k)^T f(x^k) = 0$) and the solution x^* (because the solution to the unconstrained problem satisfies $f(x^*) = 0$). Hence the point x^{k+1} is equidistant, with respect to the S norm (or equivalently, the $\hat{M}$ norm), from x^k and from x^*. Note that because the ellipses E_0^k are defined for each $k = 0, 1, 2,...$ by the same matrix $\hat{M}$, they all have the same structure and orientation. (E_0^k also has the same structure and orientation as the level sets about the solution given by $E_\alpha := \{x : (x - x^*)^T \hat{M}(x - x^*) = \alpha\}$.) Note also that the chord joining x^k to any point x on ∂E_0^k is orthogonal to the vector $f(x)$, because, by definition of E_0^k, if $x \in \partial E_0^k$, then $(x - x^k)^T f(x) = 0$. This observation reveals the relationship between the vector field defined by f and the ellipsoidal level sets.

The following result summarizes the convergence properties of the contracting ellipsoid algorithm for unconstrained problems defined by affine maps.

<u>Proposition 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 $x^* = M^{-1}b$ of $VI(f, R^n)$ from any starting point x^0 if and only if $\rho(S^{-1}M^T)$, the spectral radius of the matrix $S^{-1}M^T$, is less than one.

Moreover, for some norm $||\cdot||$, the algorithm converges linearly, with convergence ratio $||S^{-1}M^T||$.

Proof

From (8), $x^{k+1} = S^{-1}(M^T x^k + b)$, 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 $\rho(S^{-1}M^T) < 1$, then a norm $||\cdot||$ exists that satisfies $||S^{-1}M^T|| < 1$. By Cauchy's inequality, $||x^{k+1} - x^*|| = ||S^{-1}M^T(x^k - x^*)|| \leq ||S^{-1}M^T||\cdot||x^k - x^*||$ 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 (the convergence condition for the generalized steepest descent method). Consequently, if M satisfies the diagonal dominance conditions stated in Theorem 2, then the algorithm will converge. The row and column scaling procedures mentioned in Section 3 can also be used in this setting to transform the matrix M into a matrix satisfying the conditions of Proposition 1.

Lemma 1

Let M be a positive definite matrix and let $\lambda(A)$ denote the set of eigenvalues of A. Since $(M^{-1})^T = (M^T)^{-1}$, we represent this matrix by M^{-T}. Then (a) 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) $\min\{|\lambda + 1|: \lambda \in \lambda(M^{-T}M)\} > 1$; and

(vi) $\min\{|\lambda + 1|: \lambda \in \lambda(M^TM^{-1})\} > 1$;

and (b) if M^2 is positive definite, then $\rho(S^{-1}M^T) < 1$.

Proof

(a) First note that $(S^{-1}M^T)^T = MS^{-1}$, and $(M^{-T}M)^T = M^TM^{-1}$. 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 (iii) are equivalent because $(S^{-1}M^T) = (M^{-T}S)^{-1} = (M^{-T}M + I)^{-1}$.

Conditions (iii) and (v) are equivalent because

$$\rho[(M^{-T}M + I)^{-1}] = \frac{1}{\min\{|\lambda|: \lambda \in \lambda(M^{-T}M + I)\}}$$

$$= \frac{1}{\min\{|\lambda + 1|: \lambda \in \lambda(M^{-T}M)\}}$$

$$< 1$$

if and only if $\min\{|\lambda + 1|: \lambda \in \lambda(M^{-T}M)\} > 1$.

(b) 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 $\lambda \in \lambda(M^{-T}M)$, $Re(\lambda) > 0$, and hence

$$|\lambda + 1| = \{[1 + Re(\lambda)]^2 + [Im(\lambda)]^2\}^{1/2} \geq |1 + Re(\lambda)| > 1.$$

Thus, conditions (v) of part (a) holds, which ensures that (i) holds as well. ▒

Let us return to Example 1. The iterates produced by the algorithm are guaranteed to converge to the solution $x^* = \begin{bmatrix} 1/4 \\ 3/8 \end{bmatrix}$ 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, $M^2 = \begin{bmatrix} -3 & 10 \\ -10 & 12 \end{bmatrix}$ is not positive definite.

The geometrical interpretation of the algorithm discussed in Example 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. For each k, the ellipsoid E_0^k is centered about the point x^{k+1}, and $x^* \in \partial E_0^k$. In addition, the ellipsoids $\{E_0^k\}$ all have the same structure and orientation. Therefore, if $\rho(S^{-1}M^T) < 1$, then the sequence of ellipsoids converges to the point x^*, because (i) the sequence of ellipsoid centers x^{k+1} converges to x^* and (ii) x^* is on the boundary of each E_0^k.

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 $||S^{-1}M^T||_S < 1$, then the sequence of iterates x^k contracts to the solution x^* in S norm, which ensures that the ellipsoids must contract to the solution in S norm.

These observations establish the following result.

Theorem 3

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 $x^* = M^{-1}b$. Moreover, if $||S^{-1}M^T||_S < 1$, then the sequence $\{E_0^k\}$ contracts to the solution x^* in S norm.

4.2 Constrained Problems with Affine Maps

In this section, we extend the analysis of the previous section to the constrained problem VI(f,C) defined by a strictly monotone affine mapping f and a closed, convex, nonempty set $C \subseteq R^n$. We again assume that $f(x) = Mx - b$ for some positive definite nxn matrix M.

Because f is affine, the minimization subproblem (7) 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

$$(x - x^{k+1})^T[(M + M^T)x^{k+1} - (M^Tx^k + b)] \geq 0 \quad \text{for every } x \in C.$$

Hence, the subproblem is a variational inequality problem defined over C by the affine map

$$g(x,x^k) := (M + M^T)x - (M^Tx^k + b).$$

Thus, an alternative interpretation of the algorithm is that it solves a variational inequality problem defined by an asymmetric affine mapping by solving a sequence of variational inequality problems, each of which is defined by a symmetric affine mapping.

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 $S = M + M^T$. The convergence proof follows from the proof of the general contracting ellipsoid algorithm (Theorem 5).

Theorem 4

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 following example considers a constrained variational inequality problem defined by the same affine map as that in Example 1.

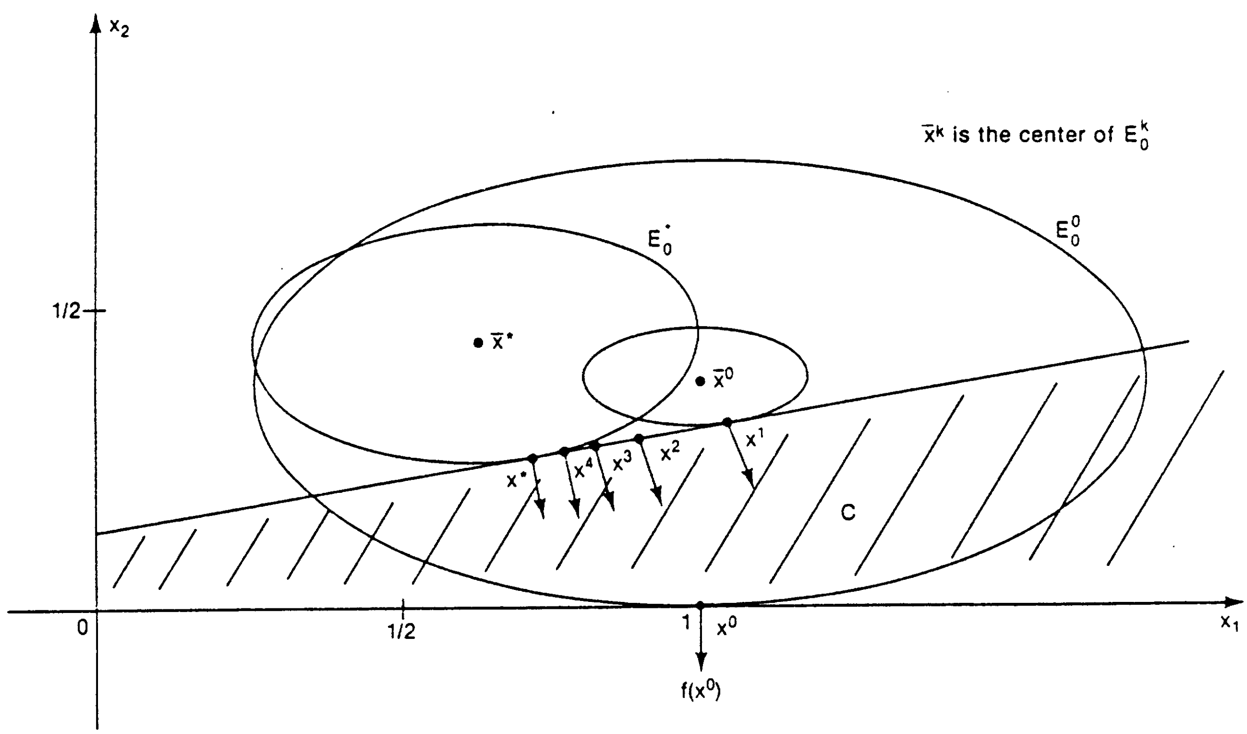
Example 2

Let $M = \begin{bmatrix} 1 & 2 \\ -2 & 4 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and let $C = \{x \in R^2: x_1 \geq 0,$ $x_2 \geq 0$, and $x_2 \leq (1/6)x_1 + 1/8\}$. The solution x^{k+1} to the k^{th} subproblem must satisfy $(x - x^{k+1})^T(Sx^{k+1} - M^Tx^k - b) \geq 0$ for every $x \in C$, where

$S = M + M^T = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$. Let $x^0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then $x^1 = \begin{bmatrix} 21/20 \\ 3/10 \end{bmatrix}$,

$x^2 = \begin{bmatrix} 9/10 \\ 11/40 \end{bmatrix}$, $x^3 = \begin{bmatrix} 33/40 \\ 21/80 \end{bmatrix}$, $x^4 = \begin{bmatrix} 63/80 \\ 41/160 \end{bmatrix}$, ...

The sequence $\{x^k\}$ converges to the solution $x^* = \begin{bmatrix} 3/4 \\ 1/4 \end{bmatrix}$ because $||S^{-1}M^T||_S = \sqrt{2}/2 < 1$. Figure 2 illustrates the sequence of iterates $\{x^k\}$ as well as a sequence of ellipses.



HBS-9084

Figure 2: The Contracting Ellipsoid Method Solves a Constrained Affine Variational Inequality Problem

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 $\bar{x}^{k+1}$ of the ellipsoid E_0^k as the next iterate x^{k+1} if $\bar{x}^{k+1}$ is a feasible point. Otherwise, it determines x^{k+1} by finding the smallest ellipsoid about $\bar{x}^{k+1}$ that contains a feasible point. This feasible point is the next iterate, x^{k+1}. This sequence of ellipsoids does not in general converge to the solution x^*. For example, in Example 2, the point $x^* = \begin{bmatrix} 3/4 \\ 1/4 \end{bmatrix}$ determines a set of ellipses centered about $\begin{bmatrix} 5/8 \\ 7/16 \end{bmatrix}$. The smallest ellipse in this set containing a feasible point contains the point x^*, thus establishing that x^* is the solution to the problem.

4.3 Constrained Problems with Nonlinear Maps

In this subsection we consider the constrained variational inequality problem VI(f,C) defined by a closed, convex subset C of R^n and a strictly monotone nonlinear map f.

For this general problem, the objective function of the minimization subproblem (7) is not necessarily convex. A solution to this subproblem must satisfy the first-order optimality conditions for problem (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, we modify the contracting ellipsoid algorithm at each iteration k by linearly approximating f about x^k. That is, we replace $f(x)$ in problem (7) with $f(x^k) + \nabla f(x^k)(x - x^k)$ to obtain the following algorithm. (Because this algorithm reduces to the Contracting Ellipsoid Method when f is affine, the proof of Theorem 4 follows from the general convergence proof of Theorem 5.)

General Contracting Ellipsoid Algorithm

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

Step 1: Let $x^{k+1} = \underset{x \varepsilon C}{\operatorname{argmin}} [(x - x^k)^T f(x^k) + (x - x^k)^T \nabla f(x^k)(x - x^k)]$.

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

Otherwise, repeat 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:

$$(x - x^{k+1})^T[[(\nabla f(x^k) + \nabla^T f(x^k)](x^{k+1} - x^k) + f(x^k)] \geq 0 \quad \text{for every } x \varepsilon C. \tag{9}$$

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.,

$$g(x, x^k) = (M_k + M_k^T)(x - x^k) + f(x^k). \tag{10}$$

Note that the general contracting ellipsoid algorithm solves a nonlinear variational inequality problem by solving a sequence of variational inequality subproblems, each of which is defined by a symmetric affine mapping.

The following theorem establishes convergence conditions for the general algorithm. We show that if $||S^{-1}M^T||_S < 1$, where $M = \nabla f(x^*)$ and $S = M + M^T$, then there is a constant $r \in [0,1)$ that satisfies $||x^{k+1} - x^*||_S \leq r||x^k - x^*||_S$. 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. We will show that if the distance with respect to the S norm from the solution x^* to a point $\bar{x} \in C$ exceeds $r||x^k - x^*||_S$, then $(x^* - \bar{x})^T g(\bar{x}, x^k) < 0$; that is, the vector $g(\bar{x}, x^k)$ points away from the point $x^* \in C$. But, if the problem map satisfied this condition, then $\bar{x}$ could not 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 $x \in C$. Therefore, the distance with respect to the S norm from x^* to x^{k+1} must be less than $r||x^k - x^*||_S$, which ensures, since $r < 1$, that the iterates contract to the solution in S norm. Figure 3 illustrates this geometrical idea. (The general structure of this proof is similar to that of Ahn and Hogan's [1982] proof of the nonlinear Jacobi method.)

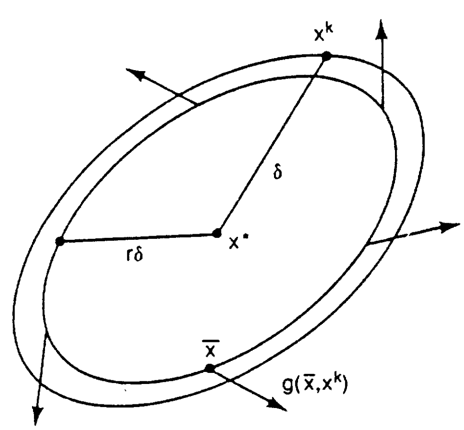


Figure 3: The Approximate Map $g(x,x^k)$ Points Away from x^* when $\| x - x^* \|_S > r \| x^k - x^* \|_S$

The convergence proof requires f to be twice 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. (If f is a strictly monotone affine map, we need not add the assumption that a solution exists: in this case existence of a solution is ensured (see Auslender [1976]) by the fact that a strictly monotone affine map is strongly monotone.)

Theorem 5

Let f be strictly monotone and twice 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 to the solution x^*, the sequence of iterates generated by the general contracting ellipsoid algorithm contracts to the solution in S norm.

Proof

We show that there is a constant $r \in [0,1)$ that satisfies

$$||x^{k+1} - x^*||_S \le r\ ||x^k - x^*||_S,$$

and hence that

$$||x^{k+1} - x^*||_S \le r^{k+1}\ ||x^0 - x^*||_S.$$

Because $r \in [0,1)$, $\lim_{k\to\infty} r^{k+1} = 0$, and hence $\lim_{k\to\infty} x^k = x^*$.

Let $K = \sup_{||x-x^*||\le 1} \left\{ \sup_{0\le t\le 1} ||S^{-1}[\nabla^2 f(x + t(x^* - x))]||_S \right\}$ and let $c = ||S^{-1}M^T||_S$. 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

$$||S^{-1}[f(x) - (f(x^*) + \nabla f(x^*)(x - x^*))]||_S \le K||x - x^*||_S^2, \tag{11}$$

and

$$||S^{-1}[\nabla f(x) - \nabla f(x^*)]||_S \le K||x - x^*||_S. \tag{12}$$

Let $\delta = ||x^k - x^*||_S$, and let $\gamma > 0$ satisfy $\gamma < \min\{\frac{1-c}{5K}, 1\}$. (Note that $\frac{1-c}{5K} > 0$, since $K > 0$ and $c < 1$.) Assume that $0 < \delta < \gamma$. Finally, let $r = (c + 3K\gamma/1-2K\gamma)$. By definition of γ, $r < 1$.

The inequality $(x - x^{k+1})^T g(x^{k+1}, x^k) \ge 0$ holds for every $x \in C$, with $g(x,x^k)$ defined by (10), because x^{k+1} solves the subproblem (9); 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 the following chain of inequalities holds:

$$\begin{aligned}
&(x^* - \bar{x})^T g(\bar{x}, x^k) \\
&\quad = (x^* - \bar{x})^T[(M_k + M_k^T)(\bar{x} - x^k) + f(x^k)] \\
&\quad = (x^* - \bar{x})^T f(x^*) + (x^* - \bar{x})^T[(M + M^T)(\bar{x} - x^*) + M^T(x^* - x^k) \\
&\qquad + M(x^* - x^k) + f(x^k) - f(x^*) + (M_k - M)(\bar{x} - x^k) \\
&\qquad + (M_k^T - M^T)(\bar{x} - x^k)] \\
&\quad < (x^* - \bar{x})^T S(\bar{x} - x^*) + (x^* - \bar{x})^T SS^{-1}M^T(x^* - x^k) \\
&\qquad + (x^* - \bar{x})SS^{-1}[M(x^* - x^k) + f(x^k) - f(x^*)] \\
&\qquad + (x^* - \bar{x})^T SS^{-1}(M_k - M)(\bar{x} - x^k) \\
&\qquad + (\bar{x} - x^k)^T SS^{-1}(M_k - M)(x^* - \bar{x}).
\end{aligned}$$

The second equality is a result of adding and subtracting terms so that we can obtain expressions in terms of the S norm; the strict inequality is valid because $(\bar{x} - x^*)^T f(x^*) > 0$.

We consider each of the terms in the last expression separately:

$$(x^* - \bar{x})^T S(\bar{x} - x^*) = -||\bar{x} - x^*||_S^2 = -(\bar{c}\delta)^2;$$

$$(x^* - \bar{x})^T S S^{-1} M^T (x^* - x^k)$$

$$= (x^* - \bar{x},\ S^{-1} M^T (x^* - x^k))_S$$

$$\leq ||x^* - \bar{x}||_S \cdot ||S^{-1} M^T||_S \cdot ||x^* - x^k||_S \qquad \text{by Cauchy's Inequality}$$

$$= \bar{c}\delta c\delta;$$

$$(x^* - \bar{x}) S S^{-1} [f(x^k) - f(x^*) - M(x^k - x^*)]$$

$$\leq ||x^* - \bar{x}||_S \cdot ||S^{-1}[f(x^k) - f(x^*) - M(x^k - x^*)]||_S$$

$$\leq ||x^* - \bar{x}||_S \cdot K||x^* - x^k||_S^2$$

$$= \bar{c}\delta K\delta^2; \qquad \text{by (11)}$$

$$(x^* - \bar{x})^T S S^{-1} (M_k - M)(\bar{x} - x^k)$$

$$\leq ||x^* - \bar{x}||_S \cdot ||S^{-1}(M_k - M)||_S \cdot ||\bar{x} - x^k||_S$$

$$\leq ||x^* - \bar{x}||_S \cdot K||x^k - x^*||_S \cdot ||\bar{x} - x^k||_S \qquad \text{by (12)}$$

$$\leq ||x^* - \bar{x}||_S \cdot K||x^k - x^*||_S (||\bar{x} - x^*||_S + ||x^* - x^k||_S)$$

by the triangle inequality

$$= \bar{c}\delta K\delta(\bar{c}\delta + \delta);$$

and similarly, $(\bar{x} - x^k)^T SS^{-1}(M_k - M)(x^* - \bar{x}) \le (\bar{c}\delta + \delta)K\delta\bar{c}\delta$.

Combining the previous inequalities, we obtain

$$(x^* - \bar{x})^T g(\bar{x}, x^k) < \bar{c}\delta^2(c + 3K\delta - \bar{c}(1 - 2K\delta))$$

$$< \bar{c}\gamma^2(c + 3K\gamma - \bar{c}(1 - 2K\gamma)). \qquad (13)$$

Now if $\bar{x} = x^{k+1}$, then as noted previously, $(x^* - \bar{x})^T g(\bar{x}, x^k) \ge 0$.
Therefore, $r = \frac{c + 3K\gamma}{1-2K\gamma} > \bar{c}$, since by (13),

$$\bar{c}\gamma^2(c + 3K\gamma - \bar{c}(1-2K\gamma)) > (x^* - \bar{x})^T g(\bar{x}, x^k) \ge 0.$$

But then,

$$||x^{k+1} - x^*||_S = \bar{c}||x^k - x^*||_S < r||x^k - x^*||_S.$$

4.4 Further Geometrical Considerations

In this subsection, we interpret the generalized steepest descent method in terms of the ellipsoidal level sets $\{E_0^k\}$ that are intrinsic to the contracting ellipsoid method, and discuss the role of symmetry in the contracting ellipsoid algorithm. We also compare convergence conditions for these two methods.

Recall that, 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. Now given x^k, the generalized steepest descent method determines the next iterate

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

with θ_k 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 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.

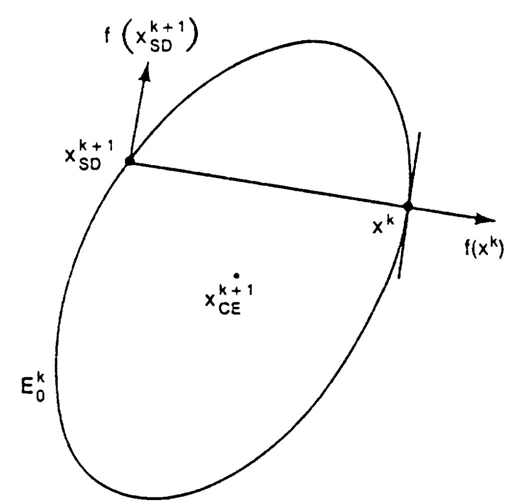


Figure 4: The Relationship Between the Ellipsoidal Level Sets and the Steepest Descent Method

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 $VI(f,R^n)$ by

$$x^{k+1} = (2M)^{-1}(Mx^k + b) = (1/2)(x^k - x^*).$$

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 5. Even if M were the identity matrix, the algorithm would still move halfway to the solution. Although in this instance the steepest descent algorithm would converge to x^* in a single iteration, in general we expect that the contracting ellipsoid algorithm would outperform the steepest descent algorithm.

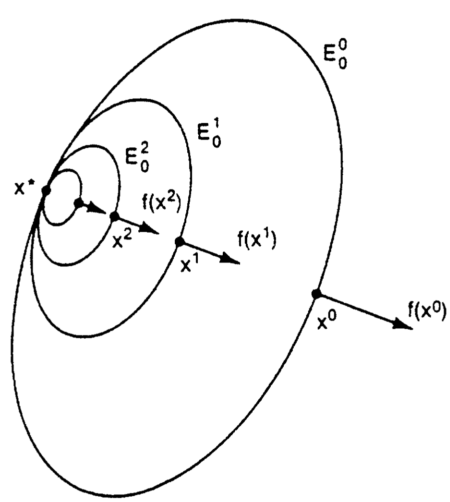


Figure 5: The Contracting Ellipsoid Iterates Move Halfway to the Solution when M is Symmetric

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 is imposed on the degree of asymmetry of M. For unconstrained problems, the generalized steepest descent algorithm is guaranteed to converge from any starting point if and only if M^2 is positive definite, while the contracting ellipsoid algorithm is guaranteed to converge from any starting point 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 1 compares these conditions for 2x2 matrices. Recall that if M^2 is positive definite, then $\rho(S^{-1}M^T) < 1$. (Hence, the steepest descent convergence conditions are more stringent than the contracting ellipsoid conditions.) Although in the 2x2 case, the conditions $\rho(S^{-1}M^T) < 1$ and $||S^{-1}M^T||_S < 1$ are identical, these conditions are not equivalent in general. (In general, $||S^{-1}M^T||_S < 1$ implies that $\rho(S^{-1}M^T) < 1$.)

4.5 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 5, we discuss the subgradient algorithm for solving a max-min problem that is equivalent to the problem VI(f,C), and show that it iteratively solves the same subproblem as the contracting ellipsoid method.

Conditions	Sample 2x2 Matrices		
	$M = \begin{bmatrix} 1 & r \\ -r & 1 \end{bmatrix}$	$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$	M symmetric $M = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$
M positive definite	True for all values of r.	$(b + c)^2 < 4ad$ $a > 0 \quad d > 0$	$ad - b^2 > 0$ $a > 0 \quad d > 0$
M^2 positive definite	$\|r\| < 1$	$(a+d)^2(b+c)^2 - 4bc(a^2+d^2) < 4a^2d^2 + 4b^2c^2$ $a^2 > -bc, \quad d^2 > -bc$	Always true if M is nonsingular.
$\rho(S^{-1}M^T) < 1$	$\|r\| < \sqrt{3}$	$b^2 + bc + c^2 < 3ad$	Always true if M is positive definite.
$\|\|S^{-1}M^T\|\|_S < 1$	$\|r\| < \sqrt{3}$	$b^2 + bc + c^2 < 3ad$	Always true if M is positive definite.

Table 1: A Comparison of Convergence Conditions for 2x2 Matrices

4.5.1 Matrix Splitting

Recall that the contracting ellipsoid method solves $VI(f,R^n)$, when $f(x) = Mx - b$, by iteratively solving the recursion (8):

$$x^{k+1} = S^{-1}(M^T x^k + b).$$

In this expression, S denotes the symmetric matrix $M + M^T$. For the unconstrained problem defined by an affine map, this algorithm is a special case of a general matrix splitting iterative method for solving linear equations. For a linear system $Mx = b$, splitting the matrix M into the sum

$$M = A - B,$$

with A chosen to be 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

$$x^{k+1} = A^{-1}(Bx^k + b) = x^k - A^{-1}(Mx^k - b).$$

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 (8) that defines the contracting ellipsoid method.

4.5.2 Projection Methods

The contracting ellipsoid method solves VI(f,C), when $f(x) = Mx - b$, by iteratively solving the subproblem (7):

$$\min \{(x - x^k)^T f(x) : x \in C\}.$$

The following lemma shows that the subproblem (7) 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 unit steplength.)

Lemma 2

If $f(x) = Mx - b$ and M is positive definite, then the subproblem

$$x^{k+1} = \underset{x \in C}{\operatorname{argmin}} \ (x - x^k)^T f(x)$$

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 operator P_C^S projecting 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

$$x^{k+1} = \underset{x \varepsilon C}{\text{argmin}} \; ||x^k - S^{-1}f(x^k) - x||_S^2$$

$$= \underset{x \varepsilon C}{\text{argmin}} \; [x^k - S^{-1}f(x^k) - x]^T S[x^k - S^{-1}f(x^k) - x]$$

$$= \underset{x \varepsilon C}{\text{argmin}} \; [Sx^k - f(x^k) - Sx]^T S^{-1}[Sx^k - f(x^k) - Sx]$$

$$= \underset{x \varepsilon C}{\text{argmin}} \; [M^T x^k + b - (M + M^T)x]^T S^{-1}[M^T x^k + b - (M + M^T)x]$$

$$= \underset{x \varepsilon C}{\text{argmin}} \; [(M^T x^k + b)^T S^{-1}(M^T x^k + b) - 2x^T(M^T x^k + b) + x^T Sx]$$

$$= \underset{x \varepsilon C}{\text{argmin}} \; 2[x^T Mx - x^T(M^T x^k + b)] \quad \text{(dropping the constant term)}$$

$$= \underset{x \varepsilon C}{\text{argmin}} \; [x^T(Mx - b) - x^T M^T x^k]$$

$$= \underset{x \varepsilon C}{\text{argmin}} \; (x - x^k)^T(Mx - b) \quad \text{(adding a constant term)}$$

$$= \underset{x \varepsilon C}{\text{argmin}} \; (x - x^k)^T f(x).$$

4.5.3 Dafermos' Framework

The contracting ellipsoid method for problems defined by affine maps fits into the framework of the general iterative scheme devised by Dafermos [1983]. 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 $x \varepsilon C$; and

(ii) $g_x(x,y)$, the partial derivative of g with respect to the first component, is symmetric and positive definite for every $x,y \varepsilon C$.

Given a point x^k, the algorithm chooses the next iterate to be the solution x^{k+1} to the following variational inequality subproblem:

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

The algorithm converges globally if g satisfies

$$\begin{aligned} &\|[g_x^{\frac{1}{2}}(x_1,y_1)]^{-T} g_y(x_2,y_2)[g_x^{\frac{1}{2}}(x_3,y_3)]^{-1}\|_2 < 1 \\ &\text{for every } x_1, x_2, x_3, y_1, y_2, y_3 \in C. \end{aligned} \tag{14}$$

Because the contracting ellipsoid method iteratively determines the point $x^{k+1} \in C$ satisfying

$$(x-x^{k+1})^T[(M + M^T)x^{k+1} - (M^Tx^k + b)] \geq 0 \quad \text{for every } x \in C,$$

the algorithm fits into the general Dafermos scheme with the association

$$g(x,y) = (M + M^T)x - (M^Ty + b).$$

Conditions (i) and (ii) are satisfied: $g(x,x) = Mx - b = f(x)$ and $g_x(x,y) = M + M^T$ is positive definite and symmetric. Because $g_y(x,y) = -M^T$, conditions (14) reduce to $\|(S^{\frac{1}{2}})^{-T}M^T(S^{\frac{1}{2}})^{-1}\|_2 < 1$. Thus, since $\|(S^{\frac{1}{2}})^{-T}M^T(S^{\frac{1}{2}})^{-1}\|_2 = \|(S^{\frac{1}{2}})^{-1}(S^{\frac{1}{2}})^{-T}M^T(S^{\frac{1}{2}})(S^{\frac{1}{2}})^{-1}\|_S = \|S^{-1}M^T\|_S$, the conditions (14) reduce to the sufficient condition for convergence specified in Theorem 4.

When f is not affine, the mapping $g(x,x^k) = \nabla^T f(x)(x - x^k) + f(x)$ defining the variational inequality subproblem of the contracting ellipsoid method is not necessarily monotone in x (as required by condition (ii) of the Dafermos method). 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 4.3 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).

For the general contracting ellipsoid method, the conditions for convergence (14) are

$$||([\nabla f(y_1) + \nabla^T f(y_1)]^{\frac{1}{2}})^{-1} \{[\nabla^2 f(y_2) + (\nabla^2 f(y_2))^T](x_2 - y_2) - \nabla f^T(y_2)\} \cdot ([\nabla f(y_3) + \nabla^T f(y_3)]^{\frac{1}{2}})^{-1}||_2 < 1$$

for every $x_1, x_2, x_3, y_1, y_2, y_3 \in C$.

These conditions are clearly much more difficult to verify than those specified in Theorem 5; namely,

$$||S^{-1}M^T||_S < 1,$$

where $M = \nabla f(x^*)$ and $S = M + M^T$, although locally they reduce to the same condition. The relationship between the contracting ellipsoid method and the Dafermos framework suggests that the types of geometrical interpretations highlighted in this paper may extend to other algorithms as well.

5. Subgradient Algorithms

In this section, we discuss the application of a subgradient algorithm (Shor [1964], Polyak [1967]; see Shapiro [1979] for a more recent exposition) to a min-max reformulation of the variational inequality problem, and show that the subgradient algorithm and the contracting ellipsoid algorithm are closely related, in that they solve the same quadratic programming subproblem.

The subgradient algorithm is usually applied to a maximization problem

$$\max \{F(x) \colon x \in C\},$$

defined by a nondifferentiable concave, continuous function F and a closed convex subset C of R^n. Given the previous iterate x^k, the algorithm determines a subgradient of F at x^k; i.e., a vector $\gamma^k \in R^n$ satisfying

$$F(x) \le F(x^k) + \gamma^k(x - x^k) \quad \text{for every} \quad x \in R^n,$$

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 denotes the projection operator onto the set C. Polyak [1969] proposes the use of a steplength α_k given by

$$\alpha_k = \lambda_k \cdot \frac{F(x^*) - F(x^k)}{||\gamma_k||^2},$$

where $0 < \varepsilon_1 \le \lambda_k \le 2 - \varepsilon_2 < 2$ and x^* maximizes F over C. He discusses several methods for choosing α_k and analyzes the convergence properties of the algorithm.

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 $f \colon R^n \to R^n$ is monotone and continuously differentiable. With these assumptions, the system of inequalities (1) is equivalent to the system of inequalities

$$(x - x^*)^T f(x) \ge 0 \quad \text{for every} \quad x \in C \tag{15}$$

(see, for example, Auslender [1976]). Thus, the problem VI(f,C) is equivalent to the max-min problem

$$\max_{x \varepsilon C} \left\{ \min_{y \varepsilon C} \{(y-x)^T f(y)\} \right\} , \tag{16}$$

or, equivalently, to the nonlinear maximization problem

$$\max \{H(x): x \varepsilon C\}, \tag{17}$$

whose objective function is given by

$$H(x) := \min_{y \varepsilon C} \{(y-x)^T f(y)\}.$$

As the pointwise minimum of functions $(y-x)^T f(y)$ that are linear in x, $H(x)$ is concave. Problem (17) is, therefore, a concave programming problem. Clearly $H(x) \leq 0$ for every $x \varepsilon C$; moreover, $H(x^*) = 0$ if and only if x^* solves VI(f,C).

The reformulation of VI(f,C) as the max-min problem (16) or (17) motivates a number of algorithms for solving VI(f,C). For example, Auslender [1976] and Nguyen and Dupuis [1984] 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 operate under the assumption 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 function H(x) is concave, and, in general, non-differentiable. Thus, the subgradient algorithm can be applied to problem (17). 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 (16).

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

$$\partial H(x) = \text{convex hull of } \{-f(\bar{y}): \bar{y} \in \underset{y \in C}{\text{argmin}} (y-x)^T f(y)\}.$$

For most problems, the application of the subgradient algorithm requires that the value $F(x^*)$ of the function at the (unknown) optimal solution be estimated at each iteration in order to specify the steplength at that iteration. For problem (17), however, this value is known to be zero. Thus, as applied to (17), the subgradient algorithm becomes:

<u>Subgradient Algorithm for VI(f,C)</u>

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

Step 1: Let $x^{k+1} = P_C[x^k - \alpha_k f(\bar{x}^k)]$,

where $\bar{x}^k = \text{argmin} \{(x-x^k)^T f(x) : x \in C\}$,

$$\alpha_k = \frac{-\lambda_k H(\bar{x}^k)}{||f(\bar{x}^k)||^2} = -\lambda_k \cdot \frac{(\bar{x}^k - x^k)^T f(\bar{x}^k)}{||f(\bar{x}^k)||^2},$$

and $0 < \varepsilon_1 \leq \lambda_k \leq 2 - \varepsilon_2 < 2$.

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 ellipsoid algorithm. However, the subgradient algorithm moves in the $-f(\bar{x}^k)$ direction from x^k, while the contracting ellipsoid algorithm moves to the point $\bar{x}^k$. (See Figure 6.)

The subgradient algorithm is particularly well-suited for solving (possibly infinite) systems of linear inequalities (Agmon [1954], Motzkin and Schoenberg [1954]). From this perspective, the algorithm works directly on the system (15) of linear inequalities (one for each $x \in C$) in the variable x^*: given x^k, the subgradient algorithm determines the point $\bar{x}^k \in C$ that defines the most violated constraint, i.e. for which $(x-x^k)^T f(x)$ for $x \in C$ is most negative, and then moves in the $-f(\bar{x}^k)$ direction.

The idea of solving VI(f,C) by moving in the direction $-f(\bar{x}^k)$ from x^k with $\bar{x}^k$ defined as

$$\bar{x}^k = \text{argmin } \{(x - x^k)^T f(x): x \in C\},$$

is reminiscent of an "extragradient" algorithm proposed by Korpelevich [1977]. This modified projection algorithm will solve variational inequality problems defined by monotone mappings. (The usual projection algorithm (see Sibony [1970], for example) requires f to be uniformly monotone). The extragradient method moves in the direction $-f(\tilde{x}^k)$ from x^k, with

$$\tilde{x}^k = P_C[x^k - \alpha f(x^k)].$$

The algorithm can be stated as follows:

Extragradient Algorithm

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 $k = k+1$.

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 contracting ellipsoid, subgradient and extragradient algorithms is more than superficial. Indeed, if $f(x) = Mx - b$, then recall from Section 4.5.1 that the solution x^{k+1} to the $(k+1)^{st}$ subproblem in the contracting ellipsoid algorithm, which equals the solution $\tilde{x}^k$ to the $(k+1)^{st}$ subproblem of the subgradient algorithm, is a projection; in fact,

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

with $S = M + M^T$. Figure 6 illustrates these three algorithms as well as the generalized steepest descent algorithm for the variational inequality problem given in Example 1, with the initial iterate given by $x^0 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$.

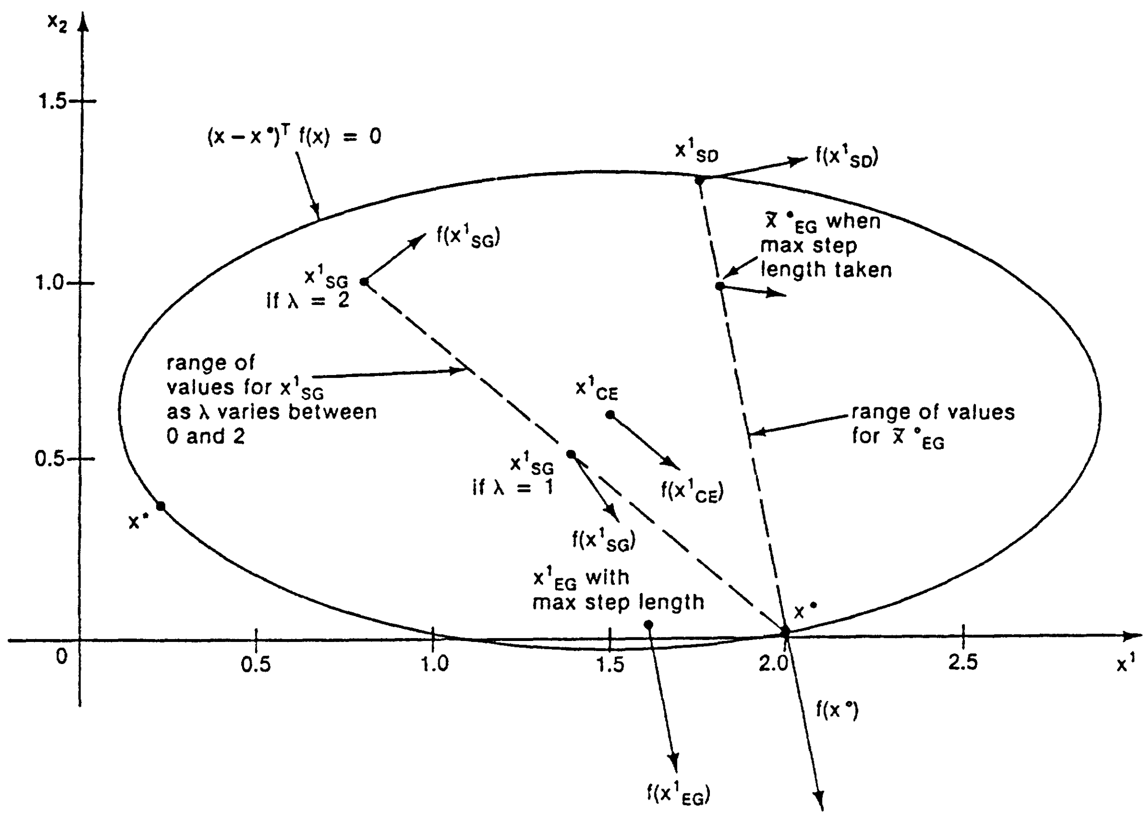


Figure 6: Geometry of Contracting Ellipsoid (CE), Steepest Descent (SD), Subgradient (SG) and Extragradient (EG) Algorithms

6. Conclusion

A monotone variational inequality problem is equivalent to a convex minimization problem whenever the Jacobian of the underlying problem map is symmetric over the feasible set. Therefore, it wouldn't be surprising if the type of conditions that allow a nonlinear programming-based algorithm to solve a variational inequality should restrict the degree of asymmetry of the underlying mapping. In this and a companion paper (Hammond and Magnanti [1985]), we formalize this notion by specifying some "near symmetry" conditions for several adaptations of nonlinear programming algorithms.

In this paper, we examine a number of algorithms for variational inequality problems that reduce to first-order approximation methods for the equivalent convex minimization problem whenever it exists. The methods that directly generalize first-order approximation methods converge under conditions (such as M^2 positive definite or $\rho((M+M^T)^{-1}M^T) < 1$) that restrict the degree of asymmetry of the problem map.

The paper focuses on the convergence of a contracting ellipsoid method. In particular, we emphasize the geometrical structure underlying the variational inequality problem that the analysis of the convergence of the contracting ellipsoid method reveals. Because this method is closely related to a number of other variational inequality algorithms, this underlying geometrical structure aids in the interpretation of those algorithms as well.

REFERENCES

Aashtiani, H. and T.L. Magnanti [1981]. "Equilibria on a Congested Transportation Network," SIAM Journal on Algebraic and Discrete Methods 2, 213-226.

Agmon, S. [1954]. "The Relaxation Method for Linear Inequalities," Canadian Journal of Mathematics 6, 382-392.

Ahn, B. and W. Hogan [1982]. "On Convergence of the PIES Algorithm for Computing Equilibria," Operations Research 30:2, 281-300.

Asmuth, R. [1978]. "Traffic Network Equilibrium," Technical Report SOL-78-2, Stanford University, Stanford, CA.

Auslender, A. [1976]. Optimisation: Méthodes Numériques, Mason, Paris.

Browder, F.E. [1966]. "Existence and Approximation of Solutions of Nonlinear Variational Inequalities," Proceedings of the National Academy of Sciences 56, 1080-1086.

Dafermos, S.C. [1980]. "Traffic Equilibrium and Variational Inequalities," Transportation Science 14, 43-54.

Dafermos, S.C. [1983]. "An Iterative Scheme for Variational Inequalities," Mathematical Programming, 26:1, 40-47.

Hammond, J.H. [1984]. "Solving Asymmetric Variational Inequality Problems and Systems of Equations with Generalized Nonlinear Programming Algorithms," Ph.D. Thesis, Department of Mathematics, M.I.T., Cambridge, MA.

Hammond, J.H. and T.L. Magnanti. [1985]. "Generalized Descent Methods for Asymmetric Systems of Equations," to appear in Mathematics of Operations Research.

Harker, P.T. [1986]. "Accelerating the Convergence of the Diagonalization and Projection Algorithms for Finite-Dimensional Variational Inequalities," Working Paper 86-03-01, Department of Decision Sciences, The Wharton School, University of Pennsylvania, Philadelphia, PA.

Hartman, P. and G. Stampacchia [1966]. "On Some Nonlinear Elliptic Differential Functional Equations," Acta Mathematica 115, 271-310.

Hearn, D.W. [1982]. "The Gap Function of a Convex Program," Operations Research Letters 1, 67-71

Khachiyan, L.G. [1979]. "A Polynomial Algorithm in Linear Programming," Soviet Mathematics Doklady, 20:1, 191-194.

Kinderlehrer, D. and G. 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.

Lions, J.L. and G. Stampacchia [1967], "Variational Inequalities," Comm. Pure Appl. Math. 20, 493-519.

Marcotte, P. [1985]. "A New Algorithm for Solving Variational Inequalities with Application to the Traffic Assignment Problem," Mathematical Programming 33, 339-351.

Motzkin, T. and I.J. Schoenberg [1954]. "The Relaxation Method for Linear Inequalities," Canadian Journal of Mathematics 6, 393-404.

Nguyen, S. and C. Dupuis. [1984]. "An Efficient Method for Computing Traffic Equilibria in Networks with Asymmetric Travel Costs," Transportation Science 18:2, 185-202.

Pang, J.-S. [1985]. "Asymmetric Variational Inequality Problems over Product Sets: Applications and Iterative Methods," Mathematical Programming 31, 206-219.

Pang, J.-S. and D. Chan [1982]. "Iterative Methods for Variational and Complementarity Problems," Mathematical Programming 24, 284-313.

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.

Shapiro, J.F. [1979]. Mathematical Programming: Structures and Algorithms, John Wiley & Sons, New York, New York.

Shor, N.Z. [1964]. "On the Structure of Algorithms for the Numerical Solution of Optimal Planning and Design Problems," Dissertation, Cybernetics Institute, Academy of Sciences, U.S.S.R.

Sibony, M. [1970]. "Méthodes Itératives pour les Equations 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.