OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Generalized Descent Methods For Asymmetric Systems of Equations and Variational Inequalities

by

Janice H. Hammond
Thomas L. Magnanti

OR 137-85 August 1985

GENERALIZED DESCENT METHODS FOR ASYMMETRIC SYSTEMS OF EQUATIONS AND VARIATIONAL INEQUALITIES

by

Janice H. Hammond
Thomas L. Magnanti

ABSTRACT

We consider generalizations of the steepest descent algorithm for solving asymmetric systems of equations. We first show that if the system is linear and is defined by a matrix M, then the method converges if M^2 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 nonlinear systems defined by uniformly monotone maps, and discuss a class of general descent methods. Finally, we show that a variant of the Frank-Wolfe method will solve a certain class of variational inequality problems.

All of the methods that we consider reduce to standard nonlinear programming algorithms for equivalent optimization problems when the Jacobian of the underlying problem map is symmetric. We interpret the convergence conditions for the generalized steepest descent algorithms as restricting the degree of asymmetry of the problem map.

KEYWORDS: variational inequalities, linear and nonlinear systems, steepest descent, Frank-Wolfe Algorithm.

This research was partially supported by Grant # ECS-8316224 from the National Science Foundation's Program in Systems Theory and Operations Research.

1. INTRODUCTION

Historically, systems of equations and inequalities have been closely linked with optimization problems. Theory and methods developed in one of these problem contexts have often complemented and stimulated new results in the other. In particular, equation and inequality systems are often viewed as the optimality conditions for an auxiliary nonlinear program. For example, in the physical sciences, variational principles (see Kaempffer [1967], for example) identify equilibrium conditions for systems such as electrical networks, chemical mixtures, and mechanical structures with equivalent optimization problems (minimizing power losses, Gibbs free energy, or potential energy). Similar identifications (e.g., identifying spatial price equilibria and urban traffic equilibria with minimizing consumer plus producer surplus (Samuelson [1952], Beckmann, McGuire and Winsten [1956])) have also proved to be quite useful in studying social and economic systems.

Once a system of equations and inequalities has been posed as an equivalent optimization problem, nonlinear programming algorithms can be used to solve the system. Indeed, many noted nonlinear programming methods have been adopted and used with considerable success to solve systems of equations (e.g., Hestenes and Stiefel [1952], Ortega and Rheinboldt [1970]).

However, viewing a system of equations or inequalities as the optimality conditions of an equivalent optimization problem requires that some form of symmetry condition be imposed on the system. For example, consider the general finite dimensional variational inequality problem VI(f,C): given a mapping $f:R^n \to R^n$ and a set $C \subseteq R^n$,

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

If f is continuously differentiable, then $f(x) = \nabla F(x)$ for some map

$F:C \subseteq R^n \to R$ if and only if $\nabla f(x)$ is symmetric for all $x \in C$. In this case the variational inequality system can be viewed as the optimality conditions for the optimization problem

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

Therefore, nonlinear programming methods can be applied to the optimization problem (2) in order to solve the variational inequality (1).

Suppose now that $\nabla f(x)$ is not symmetric, so that the variational inequality has no equivalent optimization problem such as (2). Can analogues of the nonlinear programming algorithms for (2) be applied directly to (1)? If so, when will they converge and at what convergence rate?

This paper provides partial answers to these questions. In particular, we focus on solving systems of asymmetric equations $f(x^*) = 0$, which for purposes of interpretation we view as variational inequalities with $C = R^n$ in (1). In Sections 2-5 we introduce and study a generalized steepest descent algorithm for solving the asymmetric system $f(x^*) = 0$. Section 3 considers a simplified problem setting in which f is an affine, strictly monotone mapping. Section 4 extends these results to nonlinear, uniformly monotone mappings. Section 5 shows that the convergence conditions for the generalized steepest descent method can be weakened considerably if the problem mapping is scaled in an appropriate manner. Section 6 extends the results for the generalized steepest descent method to more general gradient methods. Finally, in Section 7 we consider a generalization of the Frank-Wolfe algorithm that is applicable to constrained variational inequalities.

To conclude this section, we briefly outline the notational conventions and terminology to be used in this paper. Other definitions and notation will be introduced in the text as needed.

Let M be a real $n \times n$ matrix. In general, we define the definiteness of M without regard to symmetry: e.g., M is positive definite if and only if $x^T M x > 0$ for every nonzero $x \in R^n$. Recall that M is positive definite if and only if the symmetric part of M, defined by $\hat{M} = \frac{1}{2}(M + M^T)$, is positive definite.

An $n \times n$ positive definite symmetric matrix G defines an inner product on R^n:

$$(x,y)_G := x^T G y,$$

where $:=$ denotes definition and T denotes transposition. The inner product defined by G induces a norm on R^n:

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

which in turn induces a norm on the $n \times n$ matrix A:

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

(By writing $||A||_G$, we implicitly assume that A has the same dimensions as G.)

The mapping $f: C \subseteq R^n \to R$ is monotone on C if $(x-y)^T(f(x) - f(y)) \geq 0$ for every $x \in C$, $y \in C$; strictly monotone on C if $(x-y)^T(f(x) - f(y)) > 0$ for every $x \in C$, $y \in C$ with $x \neq y$; and uniformly (or strongly) monotone on C if 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.

Finally, 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.,

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

2. BACKGROUND AND MOTIVATION

In this section, we introduce a generalized steepest descent algorithm for asymmetric systems of equations and show that the algorithm's convergence requires some restriction on the degree of asymmetry of the problem map.

Consider the unconstrained variational inequality problem $VI(f,R^n)$, where f is continuously differentiable and uniformly monotone. This 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 is the gradient of some uniformly convex functional $F:R^n \to R$, and the unique solution x^* satisfying $f(x^*) = 0$ solves the convex minimization problem (2) with $C = R^n$. 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.

<u>Steepest Descent Algorithm for Unconstrained Minimization Problems</u>

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

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

Step 2: <u>One-Dimensional Minimization</u>. Find

$$x^{k+1} = \arg\min\{F(x) : x \in [x^k; -\nabla F(x^k)]\}.$$

Go to Step 1 with $k = k + 1$. □

Curry [1944] and Courant [1943] have given early expositions on this classical method. Curry attributes the method to Cauchy [1847], while Courant attributes it to Hadamard [1907]. As is well-known (see, for example, Polak [1971], or Bertsekas [1982]), if F is continuously differentiable and the level set $\{x : F(x) \leq F(x^0)\}$ is bounded, then the steepest descent algorithm either terminates finitely with a point x^N satisfying $\nabla F(x^N) = 0$ or it is infinite, and every limit point x^* of the sequence $\{x^k\}$ (at least one exists) satisfies $\nabla F(x^*) = 0$.

Local rate of convergence results can be obtained by approximating $F(x)$ by a quadratic function. In particular, suppose that $F(x) = ||x-x^*||_Q^2$ for some positive definite symmetric nxn matrix Q. Then if A and a are, respectively, the largest and smallest eigenvalues of Q and $r = (A-a)/(A+a)$, the sequence $\{x^k\}$ generated by the steepest descent algorithm satisfies

$$||x^{k+1} - x^*||_Q^2 = F(x^{k+1}) \leq r^2 F(x^k) = r^2 ||x^k - x^*||_Q^2. \tag{3}$$

When f is a gradient mapping, we can reformulate $VI(f,R^n)$ as the equivalent minimization problem (2) 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 $f(x) = \nabla F(x)$. 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. Therefore, the algorithm can be restated in the following equivalent form:

<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 $x^{k+1} \varepsilon [x^k;-f(x^k)]$ satisfying

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

Go to Step 1 with $k = k + 1$. □

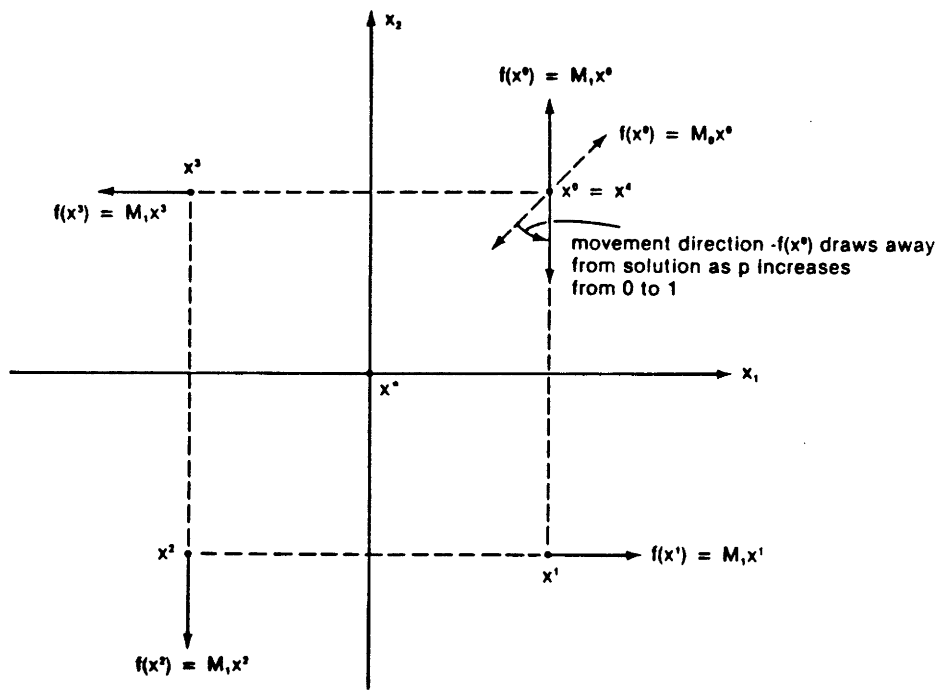
As stated, the algorithm is applicable to any unconstrained variational inequality problem. It can be viewed as a method that moves through the "vector field" defined by f by solving a sequence of one-dimensional variational inequalities.

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.

<u>Example 1</u>

Let $f(x) = M_p x$, where $x \varepsilon R^2$ and $M_p = \begin{bmatrix} 1 & -p \\ p & 1 \end{bmatrix}$. Since M_p is positive definite (because $\hat{M}_p = I$), f is uniformly monotone. If $p = 0$, f is a gradient map (since $\nabla f(x) = M_0 = I$ is symmetric) and the generalized steepest descent algorithm will converge. If $p > 0$, f is not a gradient mapping, since $\nabla f(x) = M_p$ is not symmetric. Let $x^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and consider the progress of the generalized steepest descent algorithm when $p = 1$. As long as $x^k \neq x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, 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, $-f(x^0) = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$, which implies that $x^1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, since

$f\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix}$ is orthogonal to $\begin{bmatrix}0\\-2\end{bmatrix}$. Similarly, $x^2 = \begin{bmatrix}-1\\-1\end{bmatrix}$, $x^3 = \begin{bmatrix}-1\\1\end{bmatrix}$, and $x^4 = \begin{bmatrix}1\\1\end{bmatrix} = x^0$. Thus, in this case, the algorithm cycles about the four points $\begin{bmatrix}1\\1\end{bmatrix}$, $\begin{bmatrix}1\\-1\end{bmatrix}$, $\begin{bmatrix}-1\\-1\end{bmatrix}$ and $\begin{bmatrix}-1\\1\end{bmatrix}$. Figure 1 illustrates this cyclic behavior. (In the figure, the mapping has been scaled to emphasize the orientation of the vector field.)



<u>Figure 1: The Steepest Descent Iterates Need Not Converge If M is Asymmetric</u>

The iterates produced by the generalized steepest descent algorithm do not converge when $p = 1$ because the matrix M_p is "too asymmetric": the off-diagonal entries are too large in absolute value in comparison to the diagonal entries. Geometrically, if $p = 0$, the vector field defined by f points directly away from the solution $x^* = 0$; as p increases, the vector field begins to twist, which causes the movement direction to draw away from the solution until, when $p = 1$, the algorithm no longer converges. In fact, $x^{k+1} = p\begin{bmatrix}0 & 1\\-1 & 0\end{bmatrix} x^k$, so the iterates converge to the solution if and only if $|p| < 1$.

In the following analysis, we investigate conditions on f that ensure that the generalized steepest descent algorithm solves an unconstrained variational inequality problem that cannot necessarily be reformulated as an equivalent minimization problem.

3. 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 uniformly monotone affine map. Thus, we assume that $f(x) = Mx-b$, where M is an nxn real positive definite matrix and $b \in R^n$.

3.1 Convergence of the Generalized Steepest Descent Method

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

Lemma 1

Assume that $f(x) = Mx-b$. Let x^k be the $k^{\underline{th}}$ iterate generated by the generalized steepest descent method and assume that $x^k \neq x^*$. Then the steplength determined on the $k^{\underline{th}}$ iteration is

$$\theta_k = \frac{(Mx^k-b)^T(Mx^k-b)}{(Mx^k-b)^T M(Mx^k-b)} . \qquad (4)$$

Proof

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

$$(x-x^{k+1})^T f(x^{k+1}) \geq 0 \quad \text{for every} \quad x \in [x^k; -f(x^k)]. \qquad (5)$$

If $x^{k+1} = x^k$, inequality (5) with $x = x^k - f(x^k)$ becomes $-f^T(x^k)f(x^k) \geq 0$. But then $f(x^k) = 0$, so the algorithm would have terminated in Step 1. Hence, we assume that $x^{k+1} \neq x^k$, and therefore that $\theta_k \neq 0$.

Substituting $x = x^k - \theta f(x^k)$ and $x^{k+1} = x^k - \theta_k f(x^k)$ into (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 $\theta \geq 0$, and $\theta_k > 0$, the condition $f^T(x^k)f(x^k - \theta_k f(x^k)) = 0$ must hold. Substituting $f(x) = Mx-b$ into this expression and solving for θ_k yields the expression (4). □

When f is a gradient mapping, i.e. $f(x) = \nabla F(x)$, convergence of the steepest descent algorithm follows from the fact that $F(x)$ is a descent function for the algorithm. When the Jacobian of $f(x)$ is not symmetric, no function F satisfying $\nabla F(x) = f(x)$ exists, 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 to the solution with respect to the $\hat{M}$ norm (recall that $\hat{M} = \frac{1}{2}(M + M^T)$). The $\hat{M}$ norm is a natural choice for establishing convergence because it corresponds directly to the descent function $F(x)$ when $f(x) = Mx-b$ and M is symmetric. In this case, $F(x) = (1/2)x^T Mx - b^T x$, while $||x-x^*||_M^2 = (x-x^*)^T M(x-x^*) = 2F(x) + x^{*T}Mx^*$, so $F(x)$ is a descent function for the algorithm if and only if $||x-x^*||_{\hat{M}}$ is a descent function for the algorithm.

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

Theorem 1

Let M be a positive definite matrix, and $f(x) = Mx-b$. 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

$$r = \left[1 - \inf_{x\neq 0}\left(\frac{(Mx)^T(Mx)}{x^TMx}\,\frac{x^TM^2x}{(Mx)^TM(Mx)}\right)\right]^{1/2}. \tag{6}$$

Proof

For ease of notation, let $x \neq x^*$ be the k^{th} iterate generated by the algorithm, let $\theta = \theta_k$, and let $\bar{x}$ be the $(k+1)^{st}$ iterate; then

$\bar{x} = x - \theta(Mx - b)$, where $\theta = \dfrac{(Mx-b)^T(Mx-b)}{(Mx-b)^TM(Mx-b)}$. We will show that there exists a real number $r \in [0,1)$ that is independent of x and satisfies $||\bar{x} - x^*||_{\hat{M}} \le r||x - x^*||_{\hat{M}}$. Because r must satisfy $r \ge T(x) := ||\bar{x} - x^*||_{\hat{M}}/||x - x^*||_{\hat{M}}$ 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 $r < 1$ if and only if M^2 is positive definite. Because $||z||^2_{\hat{M}} = z^TMz$ for every $z \in R^n$, we have that

$||x-x^*||_{\hat{M}} = [(x-x^*)^TM(x-x^*)]^{\frac{1}{2}}$, and

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

where the last equality follows after substituting for θ and replacing $Mx - b = M(x - M^{-1}b)$ with $M(x - x^{*})$. Thus, $T(x) = [1-R(y)]^{\frac{1}{2}}$, where $y = x-x^{*} \neq 0$ and $R(y) := \frac{[(My)^T(My)][y^TM^2y]}{[y^TMy][(My)^TM(My)]}$. Note that $r = \sup_{x \neq x^*} T(x) = \sup_{y \neq 0} [1-R(y)]^{\frac{1}{2}} = [1-\inf_{y \neq 0} R(y)]^{\frac{1}{2}} < 1$ if and only if $\inf_{y \neq 0} R(y) > 0$.

Suppose that M^2 is positive definite. Then $\widehat{M^2}$ is positive definite, and

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

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

$$= \frac{\lambda_{min}(\widehat{M^2})}{[\lambda_{max}(\hat{M})]^2} = \frac{\lambda_{min}(\widehat{M^2})}{\lambda_{max}(\hat{M}^2)}, \tag{7}$$

where $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote, respectively, the minimum and maximum eigenvalues of the real symmetric matrix A. $\widehat{M^2}$, being positive definite and symmetric, has positive real eigenvalues. Similarly, the positive definiteness of M 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 $y^TM^2y \leq 0$ for some nonzero vector y. Because M is positive definite, $(My)^TM(My) > 0$ and $y^TMy > 0$. Moreover, $y \neq 0$ ensures that $(My)^T(My) > 0$. Thus, $R(y) \leq 0$, which implies that $r \geq 1$.

The expression defining the contraction constant follows from the convergence proof. □

Corollary

The contraction constant r is bounded from above by

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

Proof

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

$$\inf_{y\neq 0} R(y) = \inf_{y\neq 0} \left\{ \frac{y^T M^2 y \cdot (My)^T(My)}{y^T My \cdot (My)^T M(My)} \right\} \geq \frac{\inf_{y\neq 0} \left\{ \frac{y^T M^2 y}{y^T My} \right\}}{\sup_{y\neq 0} \left\{ \frac{(My)^T M(My)}{(My)^T(My)} \right\}} . \tag{8}$$

The numerator of the last expression can be rewritten as follows:

$$\inf_{y\neq 0} \left\{ \frac{y^T M^2 y}{y^T My} \right\} = \inf_{y\neq 0} \left\{ \frac{y^T \widehat{M^2} y}{y^T \hat{M} y} \right\}$$

$$= \inf_{y\neq 0} \left\{ \frac{y^T \widehat{M^2} y}{(\hat{M}^{\frac{1}{2}} y)^T (\hat{M}^{\frac{1}{2}} y)} \right\}$$

$$= \inf_{z\neq 0} \left\{ \frac{z^T (\hat{M}^{\frac{1}{2}})^{-T} \widehat{M^2} (\hat{M}^{\frac{1}{2}})^{-1} z}{z^T z} \right\}$$

$$= \lambda_{min}(A),$$

where $A = (\hat{M}^{\frac{1}{2}})^{-T}\widehat{M^2}(\hat{M}^{\frac{1}{2}})^{-1}$ and $\hat{M}^{\frac{1}{2}}$ is any matrix satisfying $(\hat{M}^{\frac{1}{2}})^T(\hat{M}^{\frac{1}{2}}) = \hat{M}$. Now for any nxn matrix G, λ is an eigenvalue of A if and only if λ is an eigenvalue of $G^{-1}AG$, (see, for example, Strang [1976]). In particular, if $G = \hat{M}^{\frac{1}{2}}$, then $G^{-1}AG = \hat{M}^{-1}\widehat{M^2}$, and hence

$$\lambda_{min}(A) = \lambda_{min}(\hat{M}^{-1}\widehat{M^2}). \qquad (9)$$

Finally,

$$\sup_{y \neq 0} \left\{ \frac{(My)^T M(My)}{(My)^T(My)} \right\} = \sup_{z \neq 0} \left\{ \frac{z^T \hat{M} z}{z^T z} \right\} = \lambda_{max}(\hat{M}). \qquad (10)$$

The result follows from (8), (9) and (10). □

Note that a different upper bound on r can be derived from inequality (7) in the proof of Theorem 1:

$$r \leq \tilde{r} := \left[1 - \frac{\lambda_{min}(\widehat{M^2})}{[\lambda_{max}(\hat{M})]^2} \right]^{\frac{1}{2}} = \left[1 - \frac{\lambda_{min}(\widehat{M^2})}{\lambda_{max}(\hat{M^2})} \right]^{\frac{1}{2}} .$$

In general, this bound is not as tight as the bound $\bar{r}$ given in the corollary. To see this, let S and T be symmetric matrices, and assume that $S^{-1}T$ has positive real eigenvalues. Then

$$\begin{aligned} \lambda_{min}(T) &= \min_{||x||=1} x^T T x \\ &\leq \tilde{x}^T S S^{-1} T \tilde{x} \\ &= \lambda_{min}(S^{-1}T) \cdot \tilde{x}^T S \tilde{x} \\ &\leq \lambda_{min}(S^{-1}T) \cdot \max_{||x||=1} x^T S x \\ &= \lambda_{min}(S^{-1}T)\lambda_{max}(S), \end{aligned}$$

where $\tilde{x}$ is the eigenvector corresponding to the minimum eigenvalue of $S^{-1}T$. Let $S = \hat{M}$ and $T = \widehat{M^2}$. Then $S^{-1}T = \hat{M}^{-1}\widehat{M^2}$ has positive real eigenvalues because it has the same set of eigenvalues as the positive definite symmetric matrix $(\hat{M}^{\frac{1}{2}})^{-T}\widehat{M^2}(\hat{M}^{\frac{1}{2}})^{-1}$. Thus,

$$\frac{\lambda_{min}(\widehat{M^2})}{\lambda_{max}(\hat{M})} \le \lambda_{min}(\hat{M}^{-1}\widehat{M^2}),$$

which implies that $\bar{r} = \left[1 - \frac{\lambda_{min}(\hat{M}^{-1}\widehat{M^2})}{\lambda_{max}(\hat{M})}\right]^{\frac{1}{2}} \le \left[1 - \frac{\lambda_{min}(\widehat{M^2})}{[\lambda_{max}(\hat{M})]^2}\right]^{\frac{1}{2}} = \tilde{r}.$

3.2 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: $M^2 = M^TM$ 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 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 for every nonzero vector x, the angle between the vectors M^Tx and Mx is acute.

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 $||M-M^T||/||M+M^T||$ must be less than 1. This result follows from the following proposition.

<u>Proposition 1</u> (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 $x \neq 0$.

<u>Proof</u>

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

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

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

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

Consequently, $||M-M^T||/||M+M^T|| < 1$.

For related results on the positive definiteness of M^2, see Johnson's [1972] study of complex matrices whose hermitian part is positive definite (see also Ballantine and Johnson [1975]). This thesis and subsequent paper describe conditions under which the hermitian part of the square of such a matrix is positive definite.

3.3 Discussion of the Bound on the Contraction Constant

Let us return to the problem defined in Example 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 $\widehat{M_p^2} = \begin{bmatrix} 1-p^2 & 0 \\ 0 & 1-p^2 \end{bmatrix} = (1-p^2)I$. Thus, as we would expect from the observations after the example, M_p^2 is positive definite if and only if $|p| < 1$. Moreover, $||M_p - M_p^T||_2 / ||M_p + M_p^T||_2 < 1$ if and only if M_p^2 is positive definite, since

$$\frac{||M_p - M_p^T||_2}{||M_p + M_p^T||_2} = \frac{\left|\left|\begin{matrix} 0 & -2p \\ 2p & 0 \end{matrix}\right|\right|_2}{\left|\left|\begin{matrix} 2 & 0 \\ 0 & 2 \end{matrix}\right|\right|_2} = |p|.$$

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,

$$||x^{k+1} - x^*||_{\hat{M}_p} = ||x^{k+1}||_2 = (p^2(x^k)^T I x^k)^{\frac{1}{2}}$$

$$= |p| \cdot ||x^k||_2 = |p| \cdot ||x^k - x^*||_{\hat{M}_p};$$

hence the contraction constant for the problem is $|p|$. The bound given by the corollary is also $|p|$, because $\lambda_{min}(\hat{M}^{-1}\widehat{M^2}) = \lambda_{min}(\widehat{M^2}) = 1 - p^2$ and $\lambda_{max}(\hat{M}) = 1$, giving $\bar{r} = [1-(1-p^2)]^{\frac{1}{2}} = |p|$. □

For affine problems defined by 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

$$r_s = \frac{\lambda_{max}(M) - \lambda_{min}(M)}{\lambda_{max}(M) + \lambda_{min}(M)}$$. 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 (and hence is unitarily equivalent to a diagonal matrix) 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 and use the Kantorovich inequality to obtain the upper bound r_s on the contraction constant r.

3.4 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 (i.e., row and column) diagonal dominance condition, a necessary and sufficient condition for convergence for the problem in Example 1, is not in general a sufficiently strong condition for M^2 to be positive definite.

Example 2

Let $M = \begin{bmatrix} 2 & 0 & 0 \\ .99 & 1 & 0 \\ .99 & 0 & 1 \end{bmatrix}$. Then $M^2 = \begin{bmatrix} 4 & 0 & 0 \\ 2.97 & 1 & 0 \\ 2.97 & 0 & 1 \end{bmatrix}$ 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. □

Results by Ahn and Hogan [1982], (see also Dafermos [1983] and Florian and Spiess [1982]) imply that the norm condition $||D^{-\frac{1}{2}}BD^{-\frac{1}{2}}||_2 < 1$, where $D = \text{diag}(M)$ and $B = M-D$, ensures that the Jacobi method will solve an unconstrained variational inequality problem defined by 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 $\rho(A)$ is the spectral radius of the matrix A, because $\rho(D^{-1}B) \leq ||D^{-1}B||_D = ||D^{-\frac{1}{2}}BD^{-\frac{1}{2}}||_2$.) Pang and Chan [1981] show that if M is doubly diagonally dominant, then $||D^{-\frac{1}{2}}BD^{-\frac{1}{2}}||_2 < 1$. Example 2, therefore, also demonstrates that $||D^{-\frac{1}{2}}BD^{-\frac{1}{2}}||_2 < 1$ is not a sufficiently strong condition on M to ensure that M^2 is positive definite.

The following theorem shows that 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 2

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

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

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.

Proof

Let M_{ij} be the $(i,j)^{th}$ element of M. Then the $(i,j)^{th}$ element of M^2 is

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

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

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

i.e., that $(M_{ii})^2 > -\sum\limits_{k\neq i} M_{ik}M_{ki} + \sum\limits_{j\neq i} |M_{ii}M_{ij} + M_{ij}M_{jj} + \sum\limits_{k\neq i,j} M_{ik}M_{kj}|$, (11)

and $(M_{ii})^2 > -\sum\limits_{k\neq i} M_{ik}M_{ki} + \sum\limits_{j\neq i} |M_{ii}M_{ji} + M_{ji}M_{jj} + \sum\limits_{k\neq i,j} M_{jk}M_{ki}|$. (12)

To show that (11) holds, it is enough (by Cauchy's Inequality and the triangle inequality) to show that

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

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

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

Consequently, to show (11) is true, we show that

$$M_{ii}^2 > \sum_{k\neq i} |M_{ik}| [\sum_{j\neq k} |M_{kj}|] + M_{ii} \sum_{j\neq i} |M_{ij}| + \sum_{j\neq i} M_{jj}|M_{ij}|. \qquad (13)$$

To establish (13) (and hence (11)), we introduce the quantity t defined in the statement of the theorem. 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 of (13) is bounded from above by

$$\sum_{k\neq i} |M_{ik}|(ct) + M_{ii}(ct) + \mathrm{Max}\{M_{ii}: i=1,\ldots,n\}(ct)$$

$$< c^2t^2 + M_{ii}(ct) + (ct)\,\mathrm{Max}\{M_{ii}: i=1,\ldots,n\}$$

$$\leq c^2(M_{ii})^2 + c(M_{ii})^2 + c(M_{ii})^2 \qquad \text{by (i) and (ii).}$$

Thus, (13) holds if $(M_{ii})^2(c^2 + 2c) \leq (M_{ii})^2$, or, since $(M_{ii})^2 > 0$, if $c^2 + 2c - 1 \leq 0$, which holds if and only if $c \in [-\sqrt{2}-1, \sqrt{2}-1]$. Thus, if $c = \sqrt{2}-1$, (13), and therefore (11), must hold. Similarly, if $c = \sqrt{2}-1$, then (12) must hold. These two results establish that M^2 is doubly diagonally dominant whenever the hypotheses of the theorem are satisfied.

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 hence 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, and hence that M is positive definite. □

Because the assumption that M^2 is doubly diagonally dominant is stronger than the assumption that M^2 is positive definite, the conditions imposed on M in Theorem 2 are likely to be stronger than necessary to show that M^2 is positive definite. In a number of numerical examples, we have compared the following three conditions:

(1) the conditions of Theorem 2;

(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 2, we know that conditions (1) imply conditions (2), and 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.

The conditions that Theorem 2 imposes on the off-diagonal elements of M are the least restrictive when the diagonal elements of M are all equal. In Section 5, 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 by noting that the positive definiteness of the matrices M and M^2 is preserved under unitary transformations. As a consequence, 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.

4. 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, Auslender [1976] and Kinderlehrer and Stampacchia [1980].) The existence of a solution to $VI(f,R^n)$ is ensured if f is strongly coercive and hemicontinuous. Therefore, because uniform monotonicity implies strong coercivity, 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

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 $x \neq x^*$ be the initial iterate. We will show that if $\varepsilon := ||x-x^*|| > 0$ is sufficiently small, then the iterates generated by the algorithm contract to the solution x^*. We assume that $\varepsilon < 1$.

By Step 2 of the algorithm, $\bar{x}$, the first iterate generated by the algorithm, solves the one-dimensional variational inequality problem on the ray $[x; -f(x)]$ emanating from x in the direction $-f(x)$. The proof of Lemma 1 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 the equation

$$f^T(x)f(x - \bar{\theta}f(x)) = 0. \tag{14}$$

By Lemma 2, which follows, the value $\bar{\theta}$ satisfying (14) is unique. In order to determine an expression for $\bar{\theta}$, for any $x \in S_\varepsilon := \{x : ||x - x^*|| = \varepsilon\}$ we approximate f about x^* with a linear mapping and let v_x denote the error in this linear approximation, i.e.,

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

Substituting $M(x - \bar{\theta}f(x) - x^*) + v_{\bar{x}}$ for $f(\bar{x})$ in (14) yields the following expression for $\bar{\theta}$:

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

To show that the iterates generated by the algorithm contract to the solution in $\hat{M}$ norm, we show that there exists a real number $r \in [0,1)$ that is independent of x and satisfies $||\bar{x} - x^*|| \le r||x - x^*||$. Because r must satisfy $r \ge T(x) := ||\bar{x} - x^*||/||x - x^*||$ for every $x \in S_\varepsilon$, we define $r = \sup_{x \in S_\varepsilon} T(x)$. r is clearly nonnegative, since $T(x) > 0$ for every $x \in S_\varepsilon$.

We now show that $r < 1$. Proceeding as in the proof of Theorem 1, we have that $T(x) = [1 - R(x)]^{\frac{1}{2}}$, where

$$R(x) := \frac{\bar{\theta}f^T(x)M(x - x^*) + \bar{\theta}(x - x^*)^T Mf(x) - \bar{\theta}^2 f^T(x)Mf(x)}{(x - x^*)^T M(x - x^*)} . \tag{15}$$

Substituting for $\bar{\theta}$ in (15) and replacing $f(x)$ with $M(x - x^*) + v_x$, we have that

$$R(x) = \frac{[(x - x^*)^T M^T M(x - x^*)][(x - x^*)^T M^2(x - x^*)] - E(x)}{[(M(x - x^*) + v_x]^T M[M(x - x^*) + v_x)][(x - x^*)^T M(x - x^*)]} , \tag{16}$$

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

$$\begin{aligned} -\{v_x^T M(x - x^*) + v_{\bar{x}}^T[M(x - x^*) + v_x]\}\{(x - x^*)^T Mv_x - v_{\bar{x}}^T[M(x - x^*) + v_x]\} \\ + (x - x^*)^T M^T M(x - x^*) \cdot \{(x - x^*)^T Mv_x - v_{\bar{x}}^T[M(x - x^*) + v_x]\} \\ + (x - x^*)^T M^2(x - x^*) \cdot \{v_x^T M(x - x^*) + v_{\bar{x}}^T[M(x - x^*) + v_x]\}. \end{aligned}$$

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

$$E(x) \leq |E(x)| \leq \{||v_x||\cdot||M||\cdot||x - x^*|| + ||v_{\bar{x}}||\cdot[||M||\cdot||x - x^*|| + ||v_x||]\}^2$$
$$+ \{(||M^TM|| + ||M^2||)||x - x^*||^2\}\cdot$$
$$\{||v_x||\cdot||M||\cdot||x - x^*|| + ||v_{\bar{x}}||\cdot[||M||\cdot||x - x^*|| + ||v_x||]\}$$

$$\leq \{k_1||x - x^*||^3 + k_2||x - x^*||^2\cdot[k_3||x - x^*|| + k_4||x - x^*||^2]\}^2$$
$$+ k_5||x - x^*||^2\cdot\{k_1||x - x^*||^3 + k_2||x - x^*||\cdot$$
$$[k_3||x - x^*|| + k_4||x - x^*||^2]\}$$

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

where the second inequality follows from Lemma 3, which follows, and the third inequality holds since $||x - x^*|| < 1$ for every $x \in S_\varepsilon$. $k \geq 0$ since each $k_i \geq 0$.

Since $r = \sup_{x \in S_\varepsilon} T(x) = \sup_{x \in S_\varepsilon} [1 - R(x)]^{\frac{1}{2}} = [1 - \inf_{x \in S_\varepsilon} R(x)]^{\frac{1}{2}}$, $r < 1$ if and only if $\inf_{x \in S_\varepsilon} R(x) > 0$. We therefore show that $\inf_{x \in S_\varepsilon} R(x) > 0$.

Dividing the numerator and denominator (16) by $||x - x^*||^2[(x - x^*)^TM^TM(x - x^*)]$ (recall that $||\cdot||$ denotes the $\hat{M}$ norm) and using $-E(x) \geq -k||x - x^*||^5$, we obtain

$$\inf_{x \in S_\varepsilon} R(x) \geq \inf_{x \in S_\varepsilon} \frac{\dfrac{(x - x^*)^TM^2(x - x^*)}{(x - x^*)^TM(x - x^*)} - \dfrac{k||x - x^*||^3}{(x - x^*)^TM^TM(x - x^*)}}{\dfrac{[M(x - x^*) + v_x]^TM[M(x - x^*) + v_x]}{(x - x^*)^TM^TM(x - x^*)}}$$

$$\geq \frac{\inf\limits_{x\varepsilon S_\varepsilon} \frac{(x - x^*)^T M^2 (x - x^*)}{(x - x^*)^T M (x - x^*)} - \sup\limits_{x\varepsilon S_\varepsilon} \frac{k||x - x||^3}{(x - x^*)^T M^T M (x - x^*)}}{\sup\limits_{x\varepsilon S_\varepsilon} \frac{[M(x - x^*) + v_x]^T M [M(x - x^*) + v_x]}{(x - x^*)^T M^T M (x - x^*)}} .$$

Considering each of these three expressions separately, we have:

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

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

$$\sup\limits_{x\varepsilon S_\varepsilon} \frac{k||x - x^*||^3}{(x - x^*)^T M^T M (x - x^*)} = \sup\limits_{x\varepsilon S_\varepsilon} \frac{k||x - x^*|| \frac{(x-x^*)^T \hat{M} (x-x^*)}{(x-x^*)^T (x-x^*)}}{\frac{(x - x^*)^T M^T M (x - x^*)}{(x - x^*)^T (x - x^*)}}$$

$$\leq \frac{k\varepsilon\lambda_{max}(\hat{M})}{\lambda_{min}(M^T M)} = b\varepsilon ,$$

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

$$\sup\limits_{x\varepsilon S_\varepsilon} \frac{[M(x - x^*) + v_x]^T M [M(x - x^*) + v_x]}{(x - x^*)^T M^T M (x - x^*)}$$

$$\leq \sup\limits_{x\varepsilon S_\varepsilon} \frac{[(M(x - x^*)]^T M [M(x - x^*)]}{[M(x - x^*)]\cdot[M(x - x^*)]} + \sup\limits_{x\varepsilon S_\varepsilon} \frac{v_x^T M^2 (x - x^*) + (x - x^*)^T M^T M v_x + v_x^T M v_x}{(x - x^*)^T M^T M (x - x^*)}$$

$$\leq \lambda_{max}(\hat{M}) + \sup\limits_{x\varepsilon S_\varepsilon} \frac{||v_x||\cdot||M^2||\cdot||x - x^*|| + ||x - x^*||\cdot||M^T M||\cdot||v_x|| + ||v_x||\cdot||M||\cdot||v_x|}{(x - x^*)^T M^T M (x - x^*)}$$

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

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

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

where $c \geq 0$, and hence $a := c\lambda_{max}(\hat{M})/\lambda_{min}(M^T M) \geq 0$.

Combining these inequalities gives

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

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 $b \geq 0$. Hence, the contraction constant r is less than 1.

<u>Lemma 2</u>

If f is uniformly monotone, then, for a given $x \neq x^*$, there exists a unique $\bar{\theta} > 0$ satisfying $f^T(x)f(x - \bar{\theta}f(x)) = 0$. Moreover $\bar{\theta} \leq \frac{1}{\alpha}$, where α is the modulus of monotonicity of f.

Proof

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

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

i.e., $\bar{\theta}$ satisfies $-(\theta - \bar{\theta}) f^T(x) f(x - \bar{\theta} f(x)) \geq 0$ 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 $\alpha ||f(x)||^2$:

$$\begin{aligned}(\theta_2 - \theta_1)[g(\theta_2) - g(\theta_1)] &= (\theta_2 - \theta_1) f^T(x)[f(x - \theta_1 f(x)) - f(x - \theta_2 f(x))] \\ &= [(x - \theta_1 f(x)) - (x - \theta_2 f(x))]^T [f(x - \theta_1 f(x)) - f(x - \theta_2 f(x))] \\ &\leq \alpha ||x - \theta_1 f(x) - x + \theta_2 f(x)||^2 \\ &= [\alpha ||f(x)||^2] \, (\theta_2 - \theta_1)^2. \end{aligned} \tag{17}$$

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

To show that $\bar{\theta} \leq \frac{1}{\alpha}$, we set $\theta_2 = \bar{\theta}$ and $\theta_1 = 0$ in expression (17). Since $f^T(x) f(x - \bar{\theta} f(x)) = 0$, this substitution gives

$$\bar{\theta} f^T(x) f(x) \geq \alpha \bar{\theta}^2 f^T(x) f(x).$$

Using the fact that $\bar{\theta} > 0$ and $f(x) \neq 0$ (because $x \neq x^*$), we see that $\alpha \bar{\theta} \leq 1$. Therefore, since $\alpha > 0$, $\bar{\theta} \leq \frac{1}{\alpha}$. □

Lemma 3

Assume that f has a second Gateaux derivative on the open set $S := \{x : ||x - x^*|| < 1\}$. 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 $i = 1, 2$ and 3, there exist constants $c_i \geq 0$ that satisfy the following conditions for any $x \in S$:

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

(ii) $||\bar{x} - x^*|| \leq c_2 \, ||x - x^*||$; and

(iii) $||v_{\bar{x}}|| \leq c_3||x - x^*||^2$.

Proof

(i)
$$
\begin{aligned}
||v_x|| &= ||f(x) - M(x - x^*)|| \\
&= ||f(x) - f(x^*) - \nabla f(x^*)(x - x^*)|| \\
&\leq \left\{\sup_{0\leq t\leq 1} ||\nabla^2 f[x^* + t(x - x^*)]||\right\}||x - x^*||^2 \\
&\leq \left\{\sup_{x\in S}\left\{\sup_{0\leq t\leq 1} ||\nabla^2 f[x^* + t(x - x^*)]||\right\}\right\}||x - x^*||^2 \\
&= c_1||x - x^*||^2,
\end{aligned}
$$

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)
$$
\begin{aligned}
||\bar{x} - x^*|| &= ||x - \bar{\theta}[M(x - x^*) + v_x] - x^*|| \\
&\leq ||x - x^*|| + \frac{1}{\alpha}||M||\cdot||x - x^*|| + \frac{c_1}{\alpha}||x - x^*||^2 \\
&\leq c_2||x - x^*||.
\end{aligned}
$$

where the first inequality follows from Lemma 2 and (i), and

$c_2 := 1 + \frac{1}{\alpha}\,||M|| + \frac{c_1}{\alpha} > 0$ because $||M|| > 0$, $c_1 \geq 0$, and $\alpha > 0$.

(iii) $$\begin{aligned} ||v_{\bar{x}}|| &= ||f(\bar{x}) - M(\bar{x} - x^*)|| \\ &= ||f(\bar{x}) - f(x^*) - \nabla f(x^*)(\bar{x} - x^*)|| \\ &\le \left\{ \sup_{0\le t\le 1} ||\nabla^2 f[x^* + t(\bar{x} - x^*)]|| \right\} ||\bar{x} - x^*||^2 \\ &\le \left\{ \sup_{\{\bar{x}:||\bar{x}-x^*||\le c_2\}} \left\{ \sup_{0\le t\le 1} ||\nabla^2 f[x^* + t(\bar{x} - x^*)]|| \right\} \right\} ||\bar{x} - x^*||^2 \\ &= \bar{c}_3 ||\bar{x} - x^*||^2 \\ &\le \bar{c}_3 c_2^2 ||x - x^*||^2 = c_3 ||x - x^*||^2, \end{aligned}$$

where the last inequality follows from (ii) and $c_3 \ge 0$ because $\bar{c}_3 \ge 0$ and $c_2^2 > 0$. □

5. 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 $VI(f,R^n)$ defined by the affine mapping $f(x) = Mx-b$. We show 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 2 imposes on M.

When $f(x) = Mx-b$, the unconstrained problem $VI(f,R^n)$ is equivalent to the problem of finding a solution to the linear equation $Mx = b$. If A is a nonsingular $n \times n$ 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, suppose that M has positive diagonal entries and let $D = \text{diag}(M)$. Then D^{-1} is nonsingular, and 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 2, if for every $i = 1,2,\ldots,n$,

$$\sum_{j \neq i} |(D^{-1}M)_{ij}| < ct \quad \text{and} \quad \sum_{j \neq i} |(D^{-1}M)_{ji}| < ct, \tag{18}$$

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

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

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

(Note that all diagonal entries of $D^{-1}M$ are equal.) Conditions (18) can therefore be simplified, establishing the following result.

Theorem 4

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

$$\sum_{j \neq i} |M_{ij}| < cM_{ii} \quad \text{and} \quad \sum_{j \neq i} \frac{|M_{ji}|}{M_{jj}} < c, \tag{19}$$

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

The conditions that Theorem 4 imposes on M can be considerably less restrictive than the analogous conditions that Theorem 2 imposes on M; namely, for every $i = 1,2,\ldots,n$,

$$\sum_{j\neq i} |M_{ij}| < ct \quad \text{and} \quad \sum_{j\neq i} |M_{ji}| < ct, \tag{20}$$

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

The conditions on the row sums of M in (20) are at least as restrictive as those in (19), because $t \leq M_{ii}$ for every $i = 1,2,\ldots,n$. This is also true for the column sum conditions: because $t \leq \min\{M_{ii} : i = 1,\ldots,n\}$, the column conditions in (20) imply that $\sum_{j\neq i} |M_{ji}| < c \min\{M_{ii} : i = 1,\ldots,n\}$, and hence $\sum_{j\neq i} \frac{|M_{ji}|}{M_{jj}} \leq \sum_{j\neq i} \frac{|M_{ji}|}{\text{Min}\{M_{ii}:i=1,\ldots,n\}} < c$. The conditions specified in (20) are equivalent to those given in (19) if, and only if, all of the diagonal entries of M are identical.

The following example allows us to compare conditions (19) and (20) for the problem $VI(f,R^n)$ defined by an affine map $f(x) = Mx-b$.

Example 3

Consider the effect of scaling the matrix M defined as

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

Conditions (19) for the scaled problem reduce to the inequality

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

In contrast, conditions (20) for the unscaled problem are

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

The upper bound on $|a| + |b|$ imposed by the conditions (20) 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 (19) drive $|a|$ and $|b|$ to zero). □

Analogous results can be obtained by column-scaling the matrix M. An argument similar to our discussion of row-scaling shows that if for every $i = 1,2,\ldots,n$,

$$\sum_{j \ne i} \frac{|M_{ij}|}{M_{jj}} < c \quad \text{and} \quad \sum_{j \ne i} |M_{ji}| < cM_{ii}, \quad \text{where} \quad c = \sqrt{2} - 1, \tag{21}$$

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 could define a matrix for which the algorithm will work, even if the other does not. If we column-scale the matrix of Example 3, then conditions (21) reduce to

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

Thus, in order to obtain the least restrictive conditions on M, it is better to column scale for all positive values of N.

By using a row- or column-scaling procedure, it might 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} might be positive definite, even if M is not. The following example illustrates such a situation.

Example 4

Let $M = \begin{bmatrix} 1 & 0.5 \\ 8 & 10 \end{bmatrix}$. Neither M nor M^2 is positive definite, since $\det(\hat{M}) = -8.0625 < 0$ and $\det(\widehat{M^2}) = -1665.5625 < 0$. However, both $D^{-1}M$ and $(D^{-1}M)^2$ are positive definite, since $\det(\widehat{D^{-1}M}) = 0.5775 > 0$ and $\det(\widehat{D^{-1}M)^2} = 0.27 > 0$. Consequently, an unconstrained problem defined by $f(x) = Mx-b$, and this choice of the matrix M 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. Note that column-scaling will not produce a matrix satisfying the steepest descent convergence properties: MD^{-1} is not positive definite, since $\det(\widehat{MD^{-1}}) = -15.2 < 0$. □

These scaling procedures can also be used to transform a nonlinear mapping into one that satisfies the convergence conditions given in Theorem 3 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 $[\nabla f(x^*)D^{-1}]^2$) is positive definite,

where $D = \text{diag}[\nabla f(x^*)]$.

6. GENERALIZED DESCENT ALGORITHMS

The steepest descent algorithm for the unconstrained minimization problem

$$\text{Min } \{F(x): x \in R^n\}$$

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 (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

$$D(x^k) := \{-A_k \nabla F(x^k): A_k \text{ is an } n\text{x}n \text{ positive definite matrix}\}.$$

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 steplength) 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.

The generalized descent algorithm for the unconstrained 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: Scaling Choice. Compute the scaling matrix $A_k = A(x^0,x^1,\ldots,x^k,k)$.

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

Step 3: One-Dimensional Variational Inequality. Find $x^{k+1} \varepsilon [x^k; -A_k f(x^k)]$ satisfying

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

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 5

Let M be a positive definite matrix, and $f(x) = Mx-b$. Let $\{A_k\}$ be the sequence of positive definite symmetric matrices and $\{x^k\}$ be the sequence of iterates generated by the generalized descent algorithm applied to $VI(f,C)$. Then,

(a) the steplength θ_k determined on the k^{th} iteration of the algorithm is

$$\theta_k = \frac{(Mx^k-b)^T A_k (Mx^k-b)}{(Mx^k-b)^T A_k M A_k (Mx^k-b)} \quad ; \tag{22}$$

(b) the sequence of 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

(i) $\inf_A [\lambda_{min}(\widehat{MAM})] > 0$, and

(ii) $\inf_A [\lambda_{min}(A)] > 0$;

where the infimum is taken over all positive definite symmetric matrices A; and

(c) the contraction constant r is bounded from above by

$$r' = \left\{ 1 - \frac{\inf_A \lambda_{min}[(\hat{M}^{\frac{1}{2}})^{-T} (\widehat{MAM}) (\hat{M}^{\frac{1}{2}})^{-1}]}{\sup_A \lambda_{max}[A^{\frac{1}{2}}\hat{M}(A^{\frac{1}{2}})^T]} \right\}^{1/2} .$$

Proof

(a) For ease of notation, let x be the k^{th} iterate, A be the k^{th} scaling matrix, θ the k^{th} steplength and $\bar{x} = x - \theta Af(x)$ be the $(k+1)^{st}$ iterate generated by the algorithm. As in the proof of the generalized steepest descent method, we can assume that $\bar{x} \neq x$. (Otherwise, $f(x) = 0$, and the algorithm would have terminated in Step 2 of the k^{th} iteration.) Since $\bar{x}$ solves the unconstrained one-dimensional subproblem,

$$f^T(x)Af(\bar{x}) = (Mx-b)^T A[M(x - \theta A(Mx-b)) - b] = 0.$$

Solving the last equality for θ gives expression (22).

(b) The iterates generated by the algorithm are guaranteed to contract in $\hat{M}$ norm to the solution $x^* = M^{-1}b$ if and only if there exists a real number $r \in [0,1)$ that is independent of x and satisfies

$$||\bar{x} - x^*||_{\hat{M}} \le r||x - x^*||_{\hat{M}}.$$

Thus, we define

$$r := \sup_{A, x \ne x^*} T_A(x),$$

where

$$T_A(x) := \frac{||\bar{x} - x^*||_{\hat{M}}}{||x - x^*||_{\hat{M}}} \quad \text{for} \quad x \ne x^*.$$

As in the proof of Theorem 1, we obtain a simplified expression for $T_A(x)$:

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

where $y = x - x^*$, and

$$R_A(y) := \frac{[(My)^T A(My)][(y)^T MAMy]}{[(My)^T AMA(My)][(y)^T My]} .$$

Therefore, $r = \sup_{A, x \ne x^*} T_A(x) = \sup_{A, y \ne 0} [1 - R_A(y)]^{\frac{1}{2}} = [1 - \inf_{A, y \ne 0} R_A(y)]^{\frac{1}{2}} < 1$

if and only if $\inf_{A, y \ne 0} R_A(y) > 0$. Hence, to prove (b), we show that

$$\inf_{A, y \ne 0} R_A(y) > 0 \quad \text{if} \quad \inf_A [\lambda_{min}(A)] > 0 \quad \text{and} \quad \inf_A [\lambda_{min}(\widehat{MAM})] > 0.$$

If $\inf_A [\lambda_{min}(A)] > 0$ and $\inf_A [\lambda_{min}(\widehat{MAM})] > 0$, then

$$\inf_{A,y\neq 0} R_A(y) \geq \frac{\inf\limits_{A,y\neq 0}\left\{\frac{[(My)^T A(My)]}{(My)^T(My)}\right\} \inf\limits_{A,y\neq 0}\left\{\frac{[(y)^T MAMy]}{(y)^T y}\right\}}{\sup\limits_{A,y\neq 0}\left\{\frac{[(My)^T AM\, A(My)]}{(My)^T(My)}\right\} \sup\limits_{A,y\neq 0}\left\{\frac{[(y)^T My]}{(y)^T y}\right\}}$$

$$= \frac{\inf\limits_{A} [\lambda_{min}(A)] \cdot \inf\limits_{A} [\lambda_{min}(\widehat{MAM})]}{\sup\limits_{A} [\lambda_{max}(A\hat{M}A)] \cdot \lambda_{max}(\hat{M})} > 0,$$

because $\sup\limits_{A} [\lambda_{max}(A\hat{M}A)] > 0$ and $\lambda_{max}(\hat{M}) > 0$ by the positive definiteness of $\hat{M}$. Thus, $r < 1$.

(c) By an argument analogous to the argument in the proof of the corollary to Theorem 1,

$$\inf_{A,y\neq 0} R_A(y) \geq \frac{\inf\limits_{A,y\neq 0}\left\{\frac{(y)^T MAMy}{(y)^T My}\right\}}{\sup\limits_{A,y\neq 0}\left\{\frac{(My)^T AMA(My)}{(My)^T A(My)}\right\}}$$

$$= \frac{\inf\limits_{A} \lambda_{min}[(\hat{M}^{\frac{1}{2}})^{-T} (\widehat{MAM}) (\hat{M}^{\frac{1}{2}})^{-1}]}{\sup\limits_{A} \lambda_{max}[A^{\frac{1}{2}}\hat{M}(A^{\frac{1}{2}})^T]} .$$

The result follows from this inequality and the fact that $r = [1 - \inf\limits_{A,y\neq 0} R_A(y)]^{\frac{1}{2}}$. □

If the sequence $\{A_k\}$ of scaling matrices is chosen before the sequence of iterates $\{x^k\}$ is generated, (that is, if A_k is independent of $x^0,\ldots,x^k$) then the statement in part (b) of the theorem can be strengthened. In this case, a proof that generalizes the proof of Theorem 1 shows that the sequence $\{x^k\}$ is guaranteed to contract to the solution in $\hat{M}$ norm if and only if

$$\text{(i')} \quad \inf_{k=0,1,\ldots} [\lambda_{min}(\widehat{MA_kM})] > 0 \quad \text{and}$$

$$\text{(ii')} \quad \inf_{k=0,1,\ldots} [\lambda_{min}(A_k)] > 0.$$

When A_k is independent of $x^0,\ldots,x^k$, we can also ensure convergence of the algorithm if conditions (i') and (ii') are replaced by the conditions:

$$\text{(i'')} \quad \lim_{k\to\infty} \inf[\lambda_{min}(\widehat{MA_kM})] > 0, \quad \text{and}$$

$$\text{(ii'')} \quad \lim_{k\to\infty} \inf[\lambda_{min}(A_k)] > 0.$$

In this case, the iterates do not necessarily contract to the solution.

7. THE FRANK-WOLFE ALGORITHM

Consider the constrained variational inequality problem VI(f,C), where $f:C \subseteq R^n \to R^n$ is continuously differentiable and strictly monotone and C is a bounded polyhedron. In this constrained problem setting, how might we generalize the descent methods that we have discussed for unconstrained problems? The class of feasible direction algorithms are natural candidates to consider. In this section, we study one of these methods: the Frank-Wolfe algorithm.

If $\nabla f(x)$ is symmetric for every $x \in C$, then $f(x) = [\nabla F(x)]^T$ for some strictly convex functional $F:C \to R^1$, and the unique solution x^* to VI(f,C) solves the minimization problem (2). 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.

Recall that Frank-Wolfe algorithm (Frank and Wolfe [1956]) is a linear approximation method that iteratively approximates $F(x)$ by $F^k(x) := F(x^k) + \nabla F(x^k)(x - x^k)$. On the k^{th} iteration, the algorithm determines a vertex solution v^k to the linear program

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

and then chooses as the next iterate the point x^{k+1} that minimizes F on the line segment $[x^k, v^k]$.

Frank-Wolfe Algorithm for Linearly Constrained Convex Minimization Problems

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

Step 1: Direction Choice. Given x^k, let v^k be a vertex solution to the linear program $\min_{x \in C} x^T \nabla F(x^k)$. If $(v^k)^T \nabla F(x^k) = (x^k)^T \nabla F(x^k)$, then stop: $x^k = x^*$. Otherwise, go to Step 2.

Step 2: One-Dimensional Minimization. Let w_k solve the one dimensional minimization problem:

$$\min_{0 \leq w \leq 1} F((1-w)x^k + wv^k).$$

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 effective for 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 and one-dimensional minimization problems.

If F is pseudoconvex and continuously differentiable on the bounded polyhedron C, then (see Martos [1975], for example) the Frank-Wolfe algorithm produces a sequence $\{x^k\}$ of feasible points that 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 $f(x) = \nabla F(x)$ for every $x \in C$, we can solve $VI(f,C)$ 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 by substituting f for ∇F in Step 1 and replacing the minimization problem in Step 2 with its optimality conditions, which are necessary and sufficient because F is convex. (For other modifications of the Frank-Wolfe algorithm applicable to variational inequalities, see Lawphongpanich and Hearn [1982] and Marcotte [1983].)

<u>Generalized Frank-Wolfe Method for the Linearly Constrained Variational Inequality Problem</u>

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

Step 1: <u>Direction Choice</u>. Given x^k, let v^k be a vertex solution to the linear program $\min_{x \in C} x^T f(x^k)$. 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: <u>One-Dimensional Variational Inequality</u>. Let w_k solve the following one-dimensional variational inequality problem on the line segment $[x^k, v^k]$: Find $w_k \in [0,1]$ satisfying

$$\{[(1-w)x^k+wv^k] - [(1-w_k)x^k+w_kv^k]\}^T f[(1-w_k)x^k+w_kv^k] \geq 0$$

for every $w \in [0,1]$.

Go to Step 1 with $x^{k+1} = (1-w_k)x^k + w_kv^k$ and $k = k+1$. □

This generalization of the Frank-Wolfe method is applicable to any linearly constrained variational inequality problem. If, however, f is not a gradient mapping, the algorithm need not converge to the solution of the 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 5

Let $f(x) = Mx-b$, where $M = \begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and

$C = \{x = (x_1,x_2): x_2 \leq 1/2, \sqrt{3}\, x_1 + x_2 \geq -1, -\sqrt{3}\, x_1 + x_2 \geq -1\}$.

The solution to VI(f,C) is $x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

Let $x^0 = \begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \end{bmatrix}$. The linear program of Step 1 of the generalized Frank-Wolfe algorithm solves at $v^0 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$, and the variational inequality subproblem of Step 2 solves at $x^1 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$.

Continuing in this manner, the algorithm then generates $v^1 = \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}$,

$x^2 = \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}$, $v^2 = \begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \end{bmatrix}$, and $x^3 = \begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \end{bmatrix} = x^0$. Hence, the iterates cycle about the three points x^0, x^1 and x^2. Figure 2 illustrates this cyclic behavior. (In the figure, the mapping has been scaled to emphasize the orientation of the vector field.)

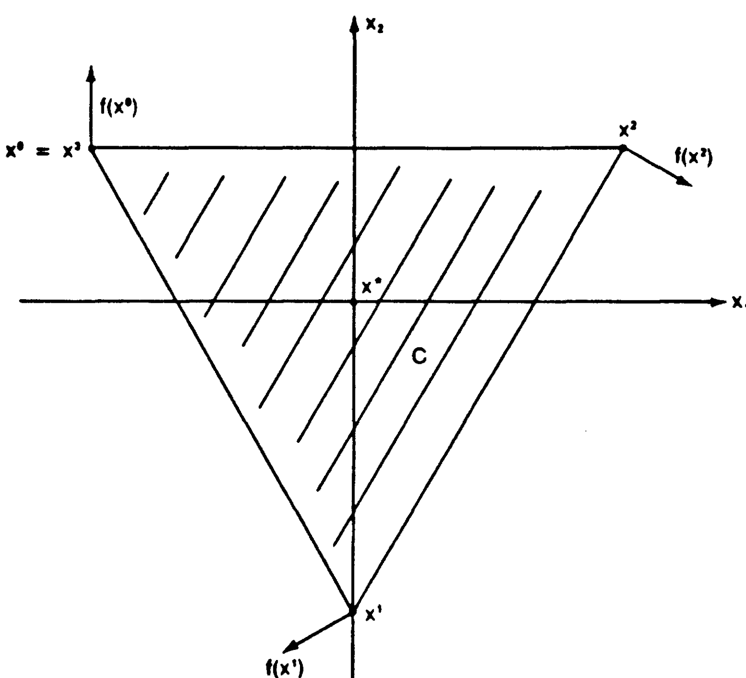


Figure 2: The Generalized Frank-Wolfe Algorithm Cycles

Example 6

Let $f(x) = Mx-b$, where $M = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ 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

$$(x - x^*)^T f(x^*) = 2/3(x_1 + x_2 + x_3 - 1) = 0 \quad \text{for every} \quad x \in C.$$

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. □

The generalized Frank-Wolfe method does not converge in the above examples because the matrix M is in some sense "too asymmetric." In all of the examples we have analyzed, the algorithm cycles only when the Jacobian of f is very asymmetric. Because the generalized Frank-Wolfe algorithm reduces to the generalized steepest descent algorithm 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 conditions 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 satisfied in neither of the previous examples. In Example 5,

$M^2 = \begin{bmatrix} -2 & 2\sqrt{3} \\ -2\sqrt{3} & -2 \end{bmatrix}$ is clearly not positive definite. In Example 6, M^2 is not positive definite because the determinant of the first 2x2 principle minor of $\widehat{M^2}$ is negative.

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. (An example in Hammond [1984] illustrates this behavior.)

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 with respect to the $\hat{M}$ norm of the unconstrained solution onto the feasible region. This is not true if M is asymmetric, so the argument cannot be generalized to the asymmetric case.

Although the Frank-Wolfe algorithm itself does not converge for either of Examples 5 or 6, the method will converge for these examples if in each iteration the step length is reduced. In particular, consider a modified version of the Frank-Wolfe method, where the step length on the k^{th} iteration is $1/k$; that is, the algorithm generates iterates by the recursion

$$x^{k+1} = x^k + \frac{1}{k+1}(v^k - x^k),$$

where v^k is the solution to the linear programming subproblem in Step 1 of the Frank-Wolfe method. This procedure can also be interpreted as an extreme-point averaging scheme: we can iteratively substitute for x^i for $i = k,k-1,\ldots,1$ to obtain

$$x^{k+1} = \frac{1}{k+1}\sum_{i=1}^{k} v^i.$$

Thus, 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 will solve the problems given in Examples 5 and 6. In the next subsection we show that it generalizes the "fictitious play" algorithm for zero-sum two-person games.

7.1 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 choosing one of the m rows of the matrix while a column player 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 row player receives $+a_{ij}$ and the column player receives $-a_{ij}$. An equilibrium solution (x^*, y^*) to the game is a pair of points $x^* \varepsilon S^m$, $y \varepsilon S^n$ satisfying

$$x^T A y^* \leq (x^*)^T A y \qquad \text{for every } x \varepsilon S^m, \; y \varepsilon S^n,$$

where S^k is the unit simplex in R^k.

The fictitious play method determines $(\bar{x}^k, \bar{y}^k)$, the k^{th} play of the game, by determining for each player the best <u>pure</u> strategy (i.e., the <u>single</u> best row or column) against the accumulated strategies of the other player. Hence, at iteration k, the row player chooses the pure strategy $\bar{x}^k$ that is the best reply 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

$$x^T A y^k \leq (\bar{x}^k)^T A y^k \qquad \text{for every } x \varepsilon S^m.$$

That is, $\bar{x}^k$ solves the (trivial) linear program

$$\max_{x \in S^m} x^T A y^k.$$

Similarly, $\bar{y}^k$ solves the linear program

$$\min_{y \in S^m} (x^k)^T A y, \qquad \text{where} \quad x^k = \frac{1}{k} \sum_{i=0}^{k-1} \bar{x}^i.$$

Robinson shows that the iterates (x^k, y^k) generated by this method converge to the equilibrium solution (x^*, y^*) of the game.

To show that the Frank-Wolfe method "with averaging" is a generalization of the fictitious play algorithm, we first reformulate the matrix game as the variational inequality problem VI(f,C), where $C = S^m x S^n$, $z = \begin{bmatrix} x \\ y \end{bmatrix}$ and $f(z) = Mz = \begin{bmatrix} 0 & -A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -Ay \\ A^T x \end{bmatrix}$. z^* solves VI(f,C) if and only if $(z-z^*)^T f(z^*) = (x-x^*)^T(-Ay^*)+(y-y^*)^T A^T x^* = (x^*)^T Ay - x^T Ay^* \geq 0$ for every $x \in S^m$, $y \in S^n$, that is, if and only if (x^*, y^*) is a solution to the game.

The Frank-Wolfe method with averaging determines an extreme point of C on the k^{th} iteration by solving the linear programming subproblem

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

(This subproblem determines $\bar{x}^k$ and $\bar{y}^k$: $\bar{z}^k$ minimizes $z^T f(z^k) = -x^T Ay^k + (x^k)^T Ay$ 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 $z^{k+1} = \frac{1}{k+1} \sum_{i=0}^{k} \bar{z}^i$: that is, $x^{k+1} = \frac{1}{k+1} \sum_{i=0}^{k} \bar{x}^i$ and $y^{k+1} = \frac{1}{k+1} \sum_{j=0}^{k} \bar{y}^j$.

Thus we can view the Frank-Wolfe algorithm with averaging as a generalized fictitious play algorithm.

The following theorem shows that this generalized 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 his example is not monotone.

Theorem 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, because v^k solves 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 $\lim_{k\to\infty} w_k = 0$. □

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 eliminates many of the applications for which the algorithm is most 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 overly 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 defined by 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).

8. CONCLUSION

In general, when nonlinear programming algorithms are adapted to variational inequality problems, their convergence requires a restriction on the degree of asymmetry of the Jacobian of the problem map. Analyzing the effect that an asymmetric Jacobian has on the vector field defined by the problem map suggests why this restriction is required. Consider the difference between the vector fields defined by two monotone affine maps, one having a symmetric Jacobian matrix and one having an asymmetric Jacobian matrix.

Let $f(x) = Mx-b$. When M is a symmetric positive definite matrix, the equation $(x-x^*)^T M(x-x^*) = c$ describes an ellipsoid whose axes are in the direction of the eigenvectors of M. The set of equations $(x-x^*)^T M(x-x^*) = c$, therefore, describe concentric ellipsoids about the solution. For any point x on the boundary of one of these ellipsoidal level sets, the vector $f(x) = Mx-b$ is normal to the hyperplane supporting the set at the point x, since $\frac{\partial}{\partial x}(x-x^*)^T M(x-x^*) = (M+M^T)(x-x^*) = 2M(x - M^{-1}b) = 2f(x)$.

If M is not symmetric, the set of equations $(x-x^*)^T M(x-x^*) = (x-x^*)^T \hat{M}(x-x^*) = c$ again describe concentric ellipsoids about the solution: the axes are in the direction of the eigenvectors of $\hat{M}$. In this instance, though, the vector $f(x) = Mx-b$ is not normal to the hyperplane supporting the ellipsoidal level set at the point x. Figure 3 illustrates the vector fields and ellipsoidal level sets for a symmetric matrix and an asymmetric matrix. In general, the more asymmetric the matrix M, the more the vector field "twists" about the origin.

Nonlinear programming algorithms are designed to solve problems defined by maps that have symmetric Jacobians. In general, these algorithms move iteratively in "good" feasible descent directions. That is, for the minimization problem (2), on the k^{th} iteration, the algorithm determines a feasible direction d_k satisfying $d_k^T \nabla F(x^k) < 0$. Many algorithms attempt to choose d_k "sufficiently close" to the steepest descent direction $-\nabla F(x^k)$. When these algorithms are adapted to solve variational inequality problems, they determine a direction d_k satisfying $d_k^T f(x^k) < 0$, with d_k close to the direction $-f(x^k)$. As long as the Jacobian of f is nearly symmetric,

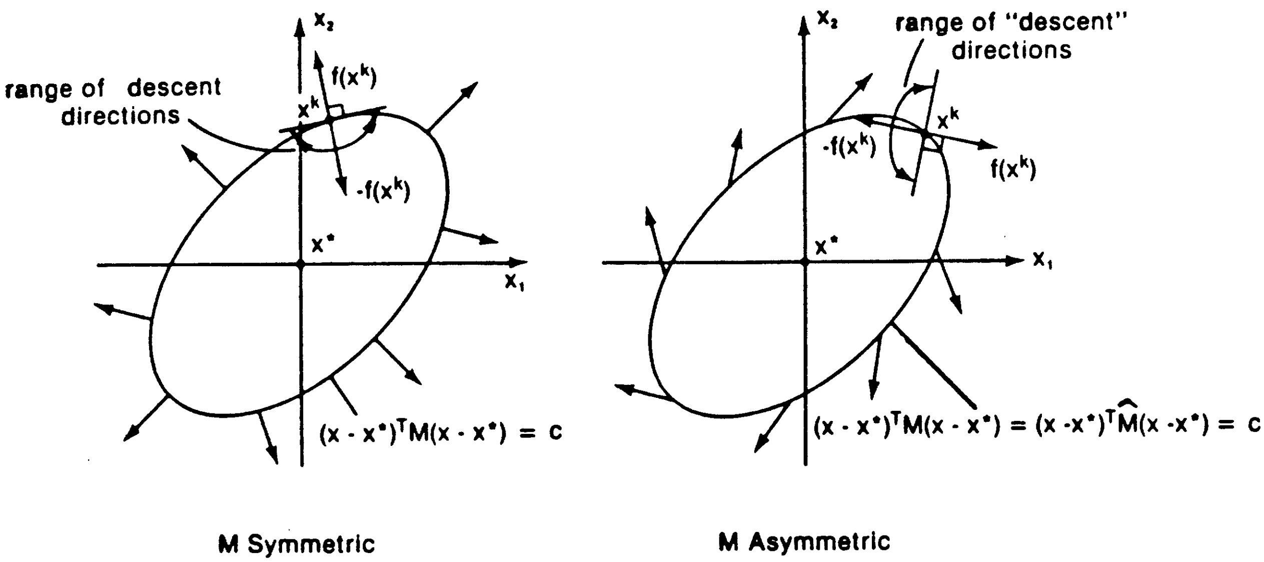


Figure 3: $f(x) = Mx-b$ is Normal to the Tangent Plane to the Ellipsoidal Level Set if and only if M is Symmetric

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 solution. 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 3 illustrates the set of "descent" directions for both the symmetric and asymmetric cases. The illustrations show that $-f(x^k)$, the direction that a nonlinear programming algorithm considers to be a "good" direction, can be a poor direction if the matrix is very asymmetric.

Projection algorithms are widely used to solve variational inequality problems. A fundamental difference between the nonlinear programming algorithms that we consider in this paper and projection methods is that the algorithms we consider 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 full steplength algorithms such as the generalized steepest descent and Frank-Wolfe algorithms require more work per iteration than 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. The asymmetry is not as much of a problem if 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 4 illustrates the effect of asymmetry on the full steplength. By our previous observations, algorithms that take a full step size will converge only if a bound is imposed on the degree of asymmetry of the Jacobian. Projection methods do not require this type of condition. The algorithms that we consider in this paper will converge even if the 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 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.

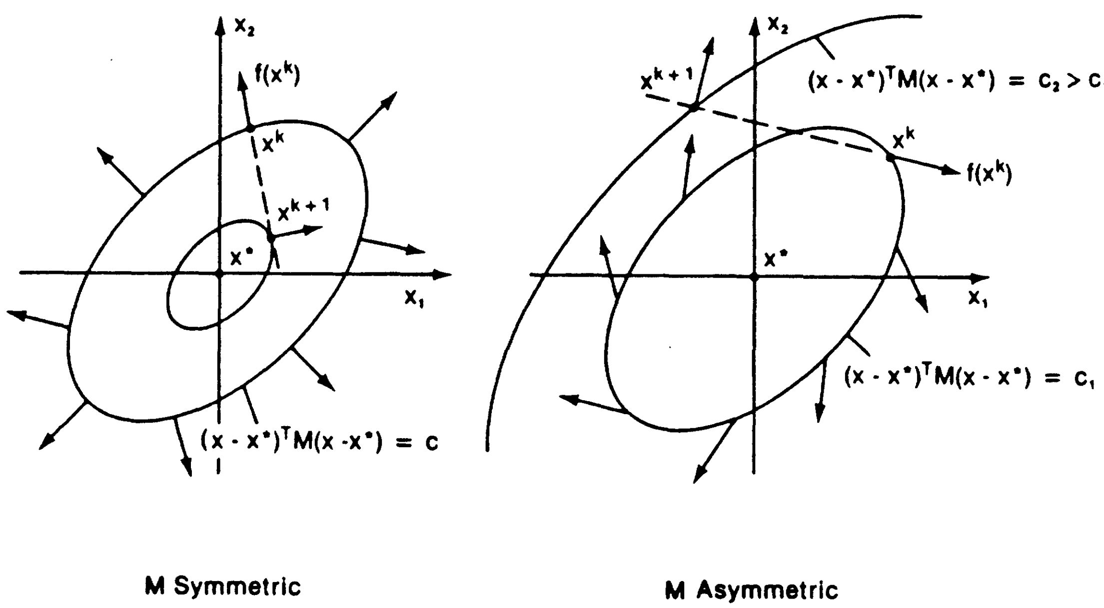


Figure 4: A Full Steplength Pulls the Iterate Further from the Solution when the Map is Very Asymmetric

REFERENCES

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.

Ballantine, C. S. and C. R. Johnson [1975]. "Accretive Matrix Products," Linear and Multilinear Algebra 3, 169-185.

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.

Bruynooghe, M., A. Gibert, and M. Sakarovitch [1968]. "Une Méthode d'Affectation du Traffic," in Fourth Symposium of the Theory of Traffic Flow, Karlsruhe.

Cauchy, A. L. [1847]. "Méthode générale pour la résolution des systèmes d'équations simultanées," Comptes rendus, Ac. Sci. Paris 25, 536-538.

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. [1983]. "An Iterative Scheme for Variational Inequlaities," Mathematical Programming, 26:1, 40-47.

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.

Gershgorin, S. [1931], "Uber 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.

Hammond, J. [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.

Hestenes, M., and E. Stiefel [1952]. "Methods of Conjugate Gradients for Solving Linear Systems," J. Res. Nat. Bur. Standards 49, 409-436.

Johnson, C. R. [1972]. "Matrices Whose Hermitian Part is Positive Definite," Ph.D. Thesis, California Institute of Technology, Pasadena, CA.

Kaempffer, F.A. [1967]. The Elements of Physics, Blaisdell Publishing Co., Waltham, MA.

Kinderlehrer, D. and Stampacchia [1980]. An Introduction to Variational Inequalities and Applications, Academic Press, New York, NY.

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.

Luenberger, D. G. [1973]. Introduction to Linear and Nonlinear Programming. Addison-Wesley, Reading, MA.

Marcotte, P. [1983]. "A New Algorithm for Solving Variational Inequalities Over Polyhedra, with Application to the Traffic Assignment Problem," Cahier du GERAD #8324, École des Hautes Étues Commerciales, Université de Montréal, Montréal, Québec.

Martos, B. [1975]. Nonlinear Programming Theory and Methods. American Elsevier Publishing Co., Inc., New York, New York.

Ortega, J. M. and W. C. Rheinboldt [1970]. Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, NY.

Pang, J. S. and D. Chan [1981]. "Iterative Methods for Variational and Complementarity Problems," Mathematical Programming 24, 284-313.

Polak, E. [19871]. Computational Methods for Optimization, Academic Press, New York, NY.

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.

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.

Strang, G. [1976]. Linear Algebra and Its Applications. Academic Press, New York, N.Y.