Averaging Schemes for Solving Fixed Point and Variational Inequality Problems

T. Magnanti and G. Perakis

OR 296-94 July 1994

Averaging Schemes for Solving Fixed Point and Variational Inequality Problems *

Thomas L. Magnanti † Georgia Perakis ‡

May 1993,
Revised July 1994

Abstract

We develop and study averaging schemes for solving fixed point and variational inequality problems. Typically, researchers have established convergence results for solution methods for these problems by establishing contractive estimates for their algorithmic maps. In this paper, we establish global convergence results using nonexpansive estimates. After first establishing convergence for a general iterative scheme for computing fixed points, we consider applications to projection and relaxation algorithms for solving variational inequality problems and to a generalized steepest descent method for solving systems of equations. As part of our development, we also establish a new interpretation of a norm condition typically used for establishing convergence of linearization schemes, by associating it with a strong-f-monotonicity condition. We conclude by applying our results to transportation networks.

*Preparation of this paper was supported, in part, by NSF Grant 9312971-DDM from the National Science Foundation.

†Sloan School of Management and Operations Research Center, MIT, Cambridge, MA 02139.

‡Operations Research Center, MIT, Cambridge, MA 02139.

1 Introduction

Fixed point and variational inequality theory provide natural frameworks for unifying the treatment of equilibrium problems and optimization encountered in problem areas as diverse as economics, game theory, transportation science, and regional science. Therefore, algorithms for computing fixed points and solving variational inequalities have widespread applicability.

The classical Banach fixed point theorem shows that for any contractive map $T(.)$ in R^n, the iterative algorithm $x_{k+1} = T(x_k)$ converges, from any starting point, to the unique fixed point of T. When the map is nonexpansive instead of contractive, this algorithm need not converge and, indeed, the map T need not have a fixed point (or it might have several). Suppose that T does have a fixed point, though, and that we "modulate" the algorithmic map by taking averages, that is, we set $x_{k+1} = \frac{x_1+T(x_1)+...T(x_k)}{k}$. Do these points converge to a fixed point? In this paper, we show that even with a more general averaging scheme, the iterates of this algorithm do converge to a fixed point of T. Using this result, we also establish, under appropriate conditions, the convergence of averages for projection and relaxation methods for solving variational inequalities and of a generalized steepest descent algorithm for solving systems of equations.

These results differ from prior averaging results in the literature. As we note in Section 2, Baillon has considered averaging of the form $x_k = \frac{y_1+...y_k}{k}$, where $y_{k+1} = T(y_k)$. That is, he averages after perfoming the iterate $y_{k+1} = T(y_k)$ rather than averaging to determine the current iterate. Bruck [8] has considered averaging for a projection algorithm applied to variational inequalities. As in Baillon's approach, if y_j is the jth iterate of the projection algorithm, he sets $x_k = \frac{a_1y_1+...+a_ky_k}{a_1+...+a_k}$. He chooses the weights a_j in a very special way; a_j is the same as the steplength for the jth iteration of the projection algorithm. His convergence results require summability conditions on the a_j rather than nonexpansiveness of the underlying map. Passty [42] has extended Bruck's results for the more general case of a forward-

backward splitting method for finding a zero of the sum of two maximal monotone operators.

The variational inequality problem (VIP) is the problem,

$$VI(f,K): \quad \text{Find} \quad x^* \in K \subseteq R^n : \; f(x^*)^t(x - x^*) \geq 0, \; \forall x \in K. \tag{1}$$

In this formulation, $f : K \subseteq R^n \rightarrow R^n$ is a given function and x^* denotes a solution of the problem. This problem is closely related to the following fixed point problem (see for example [43], [20]):

$$FP(T,K): \quad \text{Find} \quad x^* \in K \subseteq R^n \text{ satisfying } \; T(x^*) = x^*. \tag{2}$$

In this problem statement, $T : K \subseteq R^n \rightarrow K$ is a given map defined over a closed, convex (constraint) set K in R^n. The close connection between the two problems has motivated many algorithms for solving variational inequality problems ($VIPs$), such as versions of projection and relaxation algorithms. Our goal in this paper is twofold: (i) to develop and study averaging schemes for solving fixed point problems, and (ii) to apply these results to variational inequality algorithms.

The literature contains a substantial number of algorithms for the numerical solution of the variational inequality problem. The review papers of Harker and Pang [19], Pang [36] and of Florian and Hearn [13], the Ph.D. thesis of Hammond [16], the books by Harker [18] and Nagurney [33] summarize and categorize many algorithms for the problem. Pang [36] provides a survey of solution methods and underlying theory for the closely related nonlinear complementarity problem.

Projection and relaxation algorithms have a long tradition. Goldstein [15], and independently Levitin and Polyak [44], developed projection algorithms for nonlinear programming problems. Several authors, including Sibony [46], Bakusinskii and Polyak [4], Auslender [2] and Dafermos [11], have studied projection algorithms for variational inequalities while Dafermos [9], Bertsekas and Gafni [6] and others have studied these algorithms for the traffic equilibrium problem. Projection algorithms can be viewed as special cases of the linearization algorithms developed by Pang and

Chan [41]. Ahn and Hogan developed relaxation algorithms for solving economic equilibrium problems (the PIES algorithm [1]) and Dafermos considered this algorithmic approach for the general VIP, as well as the traffic equilibrium problem [11], [10]. All these algorithms are special cases of a general iterative framework developed by Dafermos [11]. Researchers originally established the convergence of projection algorithms assuming a condition of strong monotonicity on the underlying problem map f. In this paper we use a weaker condition, strong-f-monotonicity, to analyze these algorithms and establish the nonexpansiveness of their algorithmic maps (see also [14], [49], [27], [29], [43] and [31]). A problem function f is strongly-f-monotone if

$$[f(x) - f(y)]^t[x - y] \geq a\|f(x) - f(y)\|_2^2 \quad \forall x, y \in K$$

for some positive constant $a > 0$. In 1983, Gabay [14] established, in a more general context, the convergence of the projection algorithm and implicitly introduced the concept of strong-f-monotonicity. Tseng [49], using the name co-coercivity, explicitly stated this condition. Magnanti and Perakis ([27], [28], [29] and [43]) have used the term strong-f-monotonicity for this condition in order to highlight the similarity between this condition and strong monotonicity. Korpelevich [23] modified the basic projection algorithm and showed that his extragradient algorithm converges under the condition of ordinary monotonicity. Marcotte [30] tailored the extragradient method for the solution of the traffic equilibrium problem.

The steepest descent method is a standard algorithm for solving unconstrained minimization problems $\min_{\{x \in R^n\}} F(x)$, or when F is a continuously differentiable function, systems of equations of the form, find $x^* \in R^n$ satisfying $f(x^*) = \nabla F(x^*) = 0$ (see [5]). In this case, the Jacobian matrix of f is symmetric. Hammond and Magnanti [17] studied the solution of general asymmetric systems of equations, that is, find $x^* \in R^n$ satisfying $f(x^*) = 0$. They analyzed a generalized steepest descent method which generalizes the steepest descent method for the general asymmetric case (that is, when the Jacobian matrix of f is asymmetric). To obtain conver-

gence results, they require the positive definiteness of the Jacobian matrix of f and the positive definiteness of the squared Jacobian matrix of f. In this paper, using weaker assumptions, we show that averaging schemes of the generalized steepest descent method solve asymmetric systems of equations.

Another objective of this paper is to provide a better understanding of existing convergence conditions such as Pang and Chan's norm condition, and to show an "equivalency" between this norm condition (imposed on the algorithm function) and the strong-f-monotonicity condition imposed on the problem function.

The literature contains many other methods for solving variational inequalities and systems of equations. For example, Pang (see [37], [38], [39], [40]) has recently extended the work by Robinson to develop globally convergent methods (the non-smooth equation method approach) based on the solution of systems of equations which are not F-differentiable.

Researchers have typically established convergence results for many of these methods by establishing contractive estimates for their algorithmic maps $T : K \subseteq R^n \longrightarrow K$. These methods typically generate a sequence of points x_k in the feasible set K by the iterative scheme $\{x_{k+1} = T(x_k)\}_{k=0}^{\infty}$. In many cases, the convergence of this sequence to an optimal solution follows from a contraction estimate. A map T is a **contractive map on** K, relative to the $||.||_G$ norm, if

$$||T(x) - T(y)||_G \leq a||x - y||_G, \quad 0 < a < 1, \quad \forall x, y \in K.$$

In this expression, $||.||_G$ denotes the fixed norm in R^n induced by a symmetric, positive definite matrix G as $||x||_G = (x^t G x)^{1/2}$. In other cases, convergence results follow from contractive estimates only around solutions x^*, that is,

$$||T(x - T(x^*)||_G \leq a||x - x^*||_G, \quad 0 < a < 1, \quad \forall x \in K.$$

The classical Banach fixed point theorem is a standard convergence theorem for establishing the convergence of algorithms when the algorithmic map is a contraction.

In this paper, we consider nonexpansive estimates for these maps. A map T is a **nonexpansive map on** K, relative to the $||.||_G$ norm, if

$$||T(x) - T(y)||_G^2 \leq ||x - y||_G^2 \text{ for all } x, y \in K.$$

In Section 2 we develop global convergence results for a general averaging scheme induced by such nonexpansive maps. Using this result for fixed points, in Section 3 we establish the convergence of general averaging schemes for relaxation and projection algorithms for solving $VIPs$ and a generalized steepest descent method for solving general asymmetric systems of equations. Our convergence conditions are weaker than those used for establishing the convergence of these methods. We also establish an "equivalency" between the strong-f-monotonicity condition (imposed on the problem function) and the norm condition (imposed on the algorithm function) for establishing the convergence of linearization and relaxation schemes. In Section 4 we apply these results to equilibrium problems in congested transportation networks.

Finally, in Section 5, we offer some concluding remarks and raise some open questions. To conclude these introductory remarks, we review some facts concerning matrices.

Definition 1 . *A positive definite and symmetric matrix S defines an inner product $(x, y)_S = x^t S y$. The inner product induces a norm with respect to the matrix S via*

$$||x||_S^2 = x^t S x.$$

Recall that every positive definite matrix S has a square root, that is a matrix $S^{1/2}$ satisfying $S^{1/2} S^{1/2} = S$. The inner product $(x, y)_S$ is related to the Euclidean distance since

$$||x||_S = (x, x)_S^{1/2} = (x^t S x)^{1/2} = ||S^{1/2} x||_2.$$

This norm, in turn, induces an operator norm on any operator B. Namely,

$$||B||_S = \sup_{||x||_S = 1} ||Bx||_S.$$

The operator norms $||B||_S$ and $||B|| \equiv ||B||_I$ are related since

$$||B||_S = \sup_{||x||_S=1} ||Bx||_S = \sup_{||S^{1/2}x||_2=1} ||S^{1/2}Bx||_2 =$$

$$= \sup_{||S^{1/2}x||_2=1} ||S^{1/2}BS^{-1/2}S^{1/2}x||_2 = ||S^{1/2}BS^{-1/2}||_2.$$

So,

$$||B||_S = ||S^{1/2}BS^{-1/2}||$$

and, similarly,

$$||B|| = ||S^{-1/2}BS^{1/2}||_S.$$

2 Averaging schemes for solving fixed point problems

We begin by studying averaging schemes for solving fixed point problems. The analysis of these schemes rests on the nonexpansiveness of the problem maps. We also review an ergodic theorem of J.B. Baillon [3] for nonlinear nonexpansive maps which establishes the convergence of another type of an averaging scheme.

Consider a closed and convex subset K of R^n and a nonlinear map $T : K \to K$. This map could be an algorithmic map whose fixed points solve a variational inequality or, more generally, a fixed point problem of the form $FP(T, K)$. We wish to show that if the map is nonexpansive and we use the following averaging scheme

$$x_{k+1} = \frac{a_1x_1 + a_2T(x_1) + ... + a_{k+1}T(x_k)}{a_1 + ... + a_{k+1}},$$

for some appropriately chosen values of the averaging constants $a_k > 0$, then the iterates converge to a fixed point of T (assuming one exists). Setting $a(k+1) = \frac{a_{k+1}}{a_1+...+a_{k+1}}$, we could also view this scheme as an iterative method of the form $x_{k+1} = x_k + a(k+1)[T(x_k) - x_k]$. Another way to view this scheme is by setting $y_k = T(x_{k-1})$, then the induced sequence is

$$y_k = T(\frac{a_1y_1 + a_2y_2 + ... + a_{k-1}y_{k-1}}{a_1 + ... + a_{k-1}}).$$

We first prove some preliminary propositions and lemmas.

LEMMA 1:

For every fixed point x^* of the nonexpansive map T, the sequence $||x_k - x^*||_G$ is a decreasing, convergent sequence.

Proof: Since x^* is a fixed point of the map T, T is nonexpansive, and $x_{k+1} = x_k + a(k+1)[T(x_k) - x_k]$.

$$||x_{k+1} - x^*||_G \leq (1 - a(k+1))||x_k - x^*||_G + a(k+1)||T(x_k) - T(x^*)||_G \leq$$

$$(1 - a(k+1))||x_k - x^*||_G + a(k+1)||x_k - x^*|| = ||x_k - x^*||_G.$$

Therefore, $0 \leq ||x_k - x^*||_G$ is a decreasing sequence, and so it converges. Q.E.D.

LEMMA 2:

The sequence $||T(x_k) - x_k||_G$ is a decreasing, convergent sequence.

Proof: Since $(x_{k+1} - x_k) = a(k+1)[T(x_k) - x_k]$ and $x_k = x_{k-1} + a(k)[T(x_{k-1}) - x_{k-1}]$,

$$\frac{x_{k+1} - x_k}{a(k+1)} = T(x_k) + (\frac{1 - a(k)}{a(k)})x_k - \frac{1}{a(k)}x_k =$$

$$T(x_k) + (\frac{1 - a(k)}{a(k)})x_k - \frac{1}{a(k)}(x_{k-1} + a(k)[T(x_{k-1}) - x_{k-1}] =$$

$$T(x_k) - T(x_{k-1}) + (\frac{1 - a(k)}{a(k)})(x_k - x_{k-1}).$$

The nonexpansiveness of the map T and the triangle inequality imply that

$$\frac{||x_{k+1} - x_k||_G}{a(k+1)} \leq ||x_k - x_{k-1}||_G + (\frac{1 - a(k)}{a(k)})||x_k - x_{k-1}||_G = \frac{||x_k - x_{k-1}||_G}{a(k)}.$$

Since $T(x_k) - x_k = \frac{x_{k+1} - x_k}{a(k+1)}$, this result implies that $0 \leq ||T(x_k) - x_k||_G \leq ||T(x_{k-1}) - x_{k-1}||_G$, and so $||T(x_k) - x_k||_G$ is a decreasing and, therefore, a convergent sequence. Q.E.D.

Proposition 1:

Let $c_k = c_k(x^*) = \frac{||x_{k-1} - T(x_{k-1})||_G^2}{||x_{k-1} - x^*||_G^2}$. Then the following inequality is valid:

$$||x_k - x^*||_G^2 \leq e^{-\min(1, \frac{c_k}{2})\min(a(k), 1 - a(k))}||x_{k-1} - x^*||_G^2.$$

Proof: Since $x_k = x_{k-1} + a(k)[T(x_{k-1}) - x_{k-1}]$,

$$x_k - x^* = [1 - a(k)](x_{k-1} - x^*) + a(k)[T(x_{k-1}) - x^*].$$

Therefore,

$$||x_k - x^*||_G^2 = \tag{3}$$

$$a(k)^2||T(x_{k-1})-x^*||_G^2+(a(k)^2-2a(k)+1)||x_{k-1}-x^*||_G^2+2a(k)(1-a(k))(T(x_{k-1})-x^*)^tG(x_{k-1}-x^*).$$

Consider the following two cases.

1. Suppose $a(k) \leq 1/2$, then

   (a) If $(T(x_{k-1}) - x^*)^tG(x_{k-1} - x^*) \leq 0$, then expression (3) and the fact that T is a nonexpansive map imply that

$$||x_k - x^*||_G^2 \leq$$

$$a(k)^2||T(x_{k-1}) - x^*||_G^2 + (a(k)^2 - 2a(k) + 1)||x_{k-1} - x^*||_G^2 \leq$$

$$(2a(k)^2 - a(k))||x_{k-1} - x^*||_G^2 + (1 - a(k))||x_{k-1} - x^*||_G^2 \leq$$

$$(1 - a(k))||x_{k-1} - x^*||_G^2.$$

   (b) If $(T(x_{k-1}) - x_{opt})^tG(x_{k-1} - x^*) \geq 0$. We first observe that

$$-a(k)(1-a(k))||T(x_{k-1})-x_{k-1}||_G^2+a(k)[||T(x_{k-1})-x^*||_G^2-||x_{k-1}-x^*||_G^2]+||x_{k-1}-x^*||_G^2 =$$

$$a(k)^2||T(x_{k-1}) - x^*||_G^2 + (a(k)^2 - 2a(k) + 1)||x_{k-1} - x^*||_G^2+$$

$$2a(k)(1 - a(k))(T(x_{k-1}) - x^*)^tG(x_{k-1} - x^*)$$

   (by adding and subtracting x^* within the term $||T(x_{k-1}) - x_{k-1}||_G^2$ of the first expression). Therefore, since $T(.)$ is nonexpansive,

$$||x_k - x^*||_G^2 =$$

$$-a(k)(1-a(k))||T(x_{k-1})-x_{k-1}||_G^2+a(k)[||T(x_{k-1})-x^*||_G^2-||x_{k-1}-x^*||_G^2]+||x_{k-1}-x^*||_G^2 \leq$$

$$-a(k)(1 - a(k))||T(x_{k-1}) - x_{k-1}||_G^2 + ||x_{k-1} - x^*||_G^2 =$$

$$[-a(k)(1 - a(k))c_k(x^*) + 1]||x_{k-1} - x^*||_G^2 =$$

$$[\frac{c_k(x^*)}{4}[1 - 2a(k)]^2 + [1 - \frac{c_k(x^*)}{4}]]||x_{k-1} - x^*||_G^2 \leq$$

$$[\frac{c_k(x^*)}{4}[1-2a(k)]+[1-\frac{c_k(x^*)}{4}]]||x_{k-1}-x^*||_G^2 =$$

$$[1-\frac{c_k(x^*)}{2}a(k)]||x_{k-1}-x^*||_G^2,$$

since $0 < a(k) \leq 1/2$. In this case,

$$||x_k - x^*||_G^2 \leq [1-\frac{c_k(x^*)}{2}a(k)]||x_{k-1}-x^*||_G^2.$$

Combining (a) and (b), we conclude that

$$||x_k - x^*||_G^2 \leq \max[(1-a(k)),(1-\frac{c_k(x^*)}{2}a(k))]||x_{k-1}-x^*||_G^2 =$$

$$(1-\min(1,\frac{c_k(x^*)}{2})a(k))||x_{k-1}-x^*||_G^2$$

Since $1-x \leq e^{-x}$ whenever $0 < x < 1$,

$$||x_k - x^*||_G^2 \leq e^{-\min(1,\frac{c_k}{2})a(k)}||x_{k-1}-x^*||_G^2.$$

2. If $a(k) \geq 1/2$, then if we set $b(k) = 1 - a(k) \leq 1/2$, expression (3) becomes

$$||x_k - x^*||_G^2 = (b(k)^2 - 2b(k) + 1)||T(x_{k-1}) - x^*||_G^2 + \tag{4}$$

$$b(k)^2||x_{k-1}-x^*||_G^2 + 2b(k)(1-b(k))(T(x_{k-1})-x^*)^t G(x_{k-1}-x^*).$$

A similar argument as the case $a(k) \leq 1/2$, but using expression (4) in place of (3), shows that

$$||x_k - x^*||_G^2 \leq e^{-\min(1,\frac{c_k}{2})b(k)}||x_{k-1}-x^*||_G^2.$$

Combining these results, we conclude that

$$||x_k - x^*||_G^2 \leq \max[e^{-\min(1,\frac{c_k}{2})(1-a(k))}, e^{-\min(1,\frac{c_k}{2})a(k)}]||x_{k-1}-x^*||_G^2,$$

$$||x_k - x^*||_G^2 \leq e^{-\min(1,\frac{c_k}{2})\min[(1-a(k)),a(k)]}||x_{k-1}-x^*||_G^2. \quad \text{Q.E.D.}$$

Proposition 2:

The following two statements are equivalent:

1. For some fixed point x^*, $||x_{k-1} - x^*||_G^2 \longrightarrow_{k\to\infty} 0$.

2. $||x_{k-1} - T(x_{k-1})||_G^2 \longrightarrow_{k\to\infty} 0$.

Proof: "$\Rightarrow$" $||x_{k-1} - x^*||_G^2 \longrightarrow_{k\to\infty} 0$ for some solution x^*, then the nonexpansiveness of the map T implies that $||T(x_{k-1}) - x^*||_G^2 \longrightarrow_{k\to\infty} 0$; that is, $T(x_{k-1}) \longrightarrow_{k\to\infty} x^*$ and so $||x_{k-1} - T(x_{k-1})||^2 \longrightarrow_{k\to\infty} 0$.
"$\Leftarrow$" If $||x_{k-1} - T(x_{k-1})||^2 \longrightarrow_{k\to\infty} 0$, then since by Lemma 1, for any fixed point x^{**}, $||x_{k-1}||_G \leq ||x_1 - x^{**}||_G + ||x^{**}||_G$ some subsequence x_{k_j-1} converges to a point y. But since $T(x_{k_j-1})$ converges to $T(y)$ as well as to y, $T(y) = y$. Therefore, x_{k_j-1} converges to a fixed point y. Since by Lemma 1 again the entire sequence $||x_k - y||_G$ is convergent, $||x_{k-1} - y||_G^2 \longrightarrow_{k\to\infty} 0$ with $x^* = y$. Q.E.D.

Corollary 1:

If for every fixed point x^*, $||x_{k-1} - x^*||_G$ does not converge to zero, then there are positive constants $c(x^*)$ and k_0 satisfying the condition that for all $k \geq k_0$, $c_k(x^*) = \frac{||x_{k-1} - T(x_{k-1})||_G^2}{||x_{k-1} - x^*||_G^2} \geq c(x^*)$.

Equipped with these results, we can now prove the claim stated in the beginning of this section.

THEOREM 1:

Consider a map $T : K \to K$ defined on a closed, convex subset K of R^n and suppose the fixed point problem (2) it defines has a solution. Then if T is a nonexpansive map on K relative to the $||.||_G$ norm, the sequence

$$x_k = \frac{a_1 x_1 + a_2 T(x_1) + \ldots + a_k T(x_{k-1})}{a_1 + \ldots + a_k}, \quad x_1 \in K,$$

converges to a fixed point of the map T whenever each $a_k > 0$, $a(k) = \frac{a_k}{a_1 + \ldots + a_k}$ and $\sum_{k=1}^{\infty} \min(a(k), 1 - a(k)) = +\infty$. This fixed point is also the limit of the projection of the points x_k on the set of fixed points of map T.

Proof:

Proposition 1 implies that

$$||x_k - x^*||_G^2 \leq e^{-\min(1, \frac{c_k(x^*)}{2}) \min(a(k), 1-a(k))} ||x_{k-1} - x^*||_G^2,$$

with $c_k(x^*) = \frac{||x_{k-1}-T(x_{k-1})||_G^2}{||x_{k-1}-x^*||_G^2}$.

Assume that no subsequence of x_k converges to a fixed point of map T. That is, for every fixed point x^*, the sequence $||x_k - x^*||_G$ is bounded away from zero. Then the previous corollary implies for some positive constant $c(x^*)$ and for k_0 sufficiently large, $c_k(x^*) \geq c(x^*)$ for all $k \geq k_0$. Therefore, setting $q = q(x^*) = \min(1, \frac{c_k(x^*)}{2}) > 0$ shows that for all $k \geq k_0$

$$||x_k - x^*||_G^2 \leq e^{-q\min(a(k),1-a(k))}||x_{k-1} - x^*||_G^2.$$

But this inequality implies that

$$||x_k - x^*||_G^2 \leq e^{-q\sum_{k=1}^{\infty}\min(a(k),1-a(k))}||x_0 - x^*||_G^2 \longrightarrow_{k\to\infty} 0,$$

since $\sum_{k=1}^{\infty}\min(a(k), 1-a(k)) = +\infty$. This result contradicts our assumption that for every fixed point x^*, the sequence $||x_k - x^*||_G$ is bounded away from zero. Therefore, the sequence x_k converges to a fixed point x^* of T.

For every nonexpansive map, the set $F(T)$ of fixed points of T is a closed, convex set. The sequence $l_k = Pr_{F(T)}(x_k)$ converges to l^* which is the limit of the sequence x_k, since $l^* = Pr_{F(T)}(x^*) = x^*$. Q.E.D.

Corollary 2:

The sequence $x_k = \frac{a_1x_1+a_2T(x_1)+\ldots+a_kT(x_{k-1})}{a_1+\ldots+a_k}$, with $x_1 \in K$ and with each $a_k > 0$, defines a sequence $\{y_k = T(x_{k-1})\}$. Whenever $\sum_{k=1}^{\infty}\min(a(k), 1-a(k)) = +\infty$, and $a(k) = \frac{a_k}{a_1+\ldots+a_k}$, this sequence converges to the same fixed point of the map T as does the sequence x_k.

Proof: Let x^* be the fixed point of the map T to which the sequence x_k converges, as shown in Theorem 1. Then

$$||y_{k+1} - x^*||_G = ||T(x_k) - x^*||_G \leq ||x_k - x^*||_G \longrightarrow_{k\to\infty} 0.$$

Q.E.D.

Proposition 3:

If $1 > c_1 \geq a(k) \geq c_2 > 0$ infinitely often (or if $a(k)$ has a convergent subsequence

$a(k_j)$ converging to a point $1 > c > 0$), then the averaging sequence x_k, converges to a fixed point x^* of the map T.

Proof: If $1 > c_1 \geq a(k) \geq c_2 > 0$ infinitely, then

$\sum_{k=1}^{\infty} \min(a(k), 1 - a(k)) \geq \sum_{k_j=1}^{\infty} \min(c_1, 1 - c_2) = +\infty$, and so Theorem 1 implies this result.

In this case there is a simpler proof of this Proposition which does not require the use of Theorem 1. Lemma 1 implies that $||x_{k-1}||_G \leq ||x_0 - x^*||_G + ||x^*||_G$. Furthermore, since $1 > c_1 \geq a(k) \geq c_2 > 0$ infinitely often, the sequence $\{a(k)\}$ has a convergent subsequence $\{a(k_j)\}$ converging to a point $1 > c > 0$. Then since $||x_{k_j-1}||_G$ is bounded, perhaps going to a further subsequence, $x_{k_j-1} \longrightarrow_{k_j \to \infty} y$. Suppose $y \neq T(y)$, then

$$x_{k_j} \longrightarrow_{k_j \to \infty} y' = (1-c)y + cT(y).$$

$$\text{Therefore, } ||y' - x^*||_G^2 = ||(1-c)(y - x^*) + c(T(y) - T(x^*)||_G^2 <$$

$$(1-c)||y - x^*||_G^2 + c||T(y) - T(x^*||_G^2 \leq ||y - x^*||_G.$$

This result contradicts the fact that $||y - x^*||_G = ||y' - x^*||_G$, since $||x_k - x^*||_G$ is a convergent sequence. Therefore, the limit point y of the subsequence, x_{k_j-1} is a fixed point of the map T. Furthermore, since $||x_k - y||_G$ is a convergent sequence converging to zero across a subsequence, the entire sequence x_k converges to y, which we have already shown to be a fixed point of T. Q.E.D.

Baillon [3] has proposed a different type of averaging scheme for solving fixed point problems. He proved that the sequence of averages

$$\frac{x_1 + T(x_1) + ... + T^k(x_1)}{k+1}, \quad x_1 \in K,$$

converges to a fixed point of the map T. This averaging scheme differs from the one we have considered in two respects (i) each $a_k = 1$, which is a special case of our convergence condition, and (ii) this scheme considers averages of the original

sequence $z_k = T^k(x_1)$, while our averaging scheme considers the averages of the previous iterates, that is

$$y_k = T(\frac{a_1 y_1 + a_2 y_2 + ... + a_{k-1} y_{k-1}}{a_1 + ... + a_{k-1}}).$$

The following theorem summarizes Baillon's result.

THEOREM 2 (Baillon [3]):

Let T be a map, $T : K \rightarrow K$, defined on a closed, bounded and convex subset K of a Hilbert space H. If T is a nonexpansive map on K relative to the $||.||_G$ norm, then the map

$$S_k(y) = \frac{y + T(y) + ... + T^{k-1}(y)}{k}, \quad y \in K,$$

converges weakly to a fixed point of map T, which is also the strong limit of the projection of $T^k(y)$ on the set of fixed points of map T.

Remarks:

1. The averaging sequence of the form $\frac{a_1 y_1 + a_2 y_2 + ... + a_{k+1} y_{k+1}}{a_1 + ... + a_{k+1}}$ is sometimes referred to in the literature as the Riesz process [21].

2. When $a_k = 1$, $a(k) = \frac{1}{k}$. In this case, the series $\sum_{k=1}^{\infty} \min(a(k), 1 - a(k)) = +\infty$ since for the harmonic series, $\sum_{k=1}^{\infty} \frac{1}{k} = +\infty$. When $a_k = a^k$ with $a > 1$, $a(k) = \frac{a^k(a-1)}{a(a^k-1)}$. Then $a(k)$ converges to $1 - 1/a$, which is strictly less than 1 and strictly greater than 0, and Proposition 3 (or Theorem 1) implies the convergence result.

3. In Theorem 1 we did not need to assume the feasible set K to be bounded.

4. The convergence proof of Theorem 1 rests primarily on the fact that T is a nonexpansive map, with respect to the G norm, around the solutions x^*. Only in the proof of Lemma 2, where we prove the convergence of the sequence $||T(x_k) - x_k||_G$, did we also require the nonexpansiveness of map T around every feasible point. If we assume a stronger assumption on the series

$\sum_k \min(a(k), 1-a(k))$ namely, that for every subsequence $\{a(k_j)\}_{k_j \in N}$ the series $\sum_{k_j \in N} \min(a(k_j), 1-a(k_j)) = +\infty$, then we can still establish the results of Theorem 1 with the only requirement that T is a nonexpansive map, with respect to the G norm, around the solutions x^*. Letting $a_k = a^k$ with $a > 1$ defines such a sequence. Using this observation, we can obtain the convergence of averaging schemes of "Riesz" type for maps that are nonexpansive only around the fixed points. This type of result is useful because for some algorithms (for example, the generalized steepest descent method for solving $VIPs$), it is less restrictive to establish nonexpansiveness only around the fixed point solutions. Baillon's averaging scheme does not permit us to exploit this possibility.

3 On the convergence of sequences of averages for VIP algorithms

In this section we consider variational inequality problems. Computational methods for computing solutions to variational inequalities often reduce the problem to a fixed point problem either by (i) iteratively using an ("easily computable") approximation $f'(x) = g(x, y)$ of the mapping $f(x)$ around a given point (iterate) y, with g defined so that y solves $VI(f', K)$ if and only if y solves $VI(f, K)$, or (ii) modeling the variational inequality directly as a fixed point problem. In the first case, we define the image of the underlying map $T(y)$ as a point y that solves $VI(f', K)$. We consider this type model in Subsection 3.1. One direct approach models $VI(f, K)$ as a fixed point to a projection problem. Let $Pr_K^G(y)$ denote the projection of the point y onto the set K with respect to the norm induced by the positive define matrix G. If K is a convex set, then the optimality conditions of the problem $\min_{xinK}(y - x)^t G(y - x)$ imply that x^* is a fixed point of the map $T(x) = Pr_K^G(x - \rho G^{-1} f(x))$ for any constant $\rho > 0$ if and only if x^* solves the variational inequality problem $VI(f, K)$. In this section, we study the application of Theorems 1 and 2 to both

of these approaches. We show that when the problem function of a VIP satisfies a norm condition, the sequences of averages (of Section 2) induced by relaxation algorithms converge to an optimal solution x^* of the problem. Then we establish a similar result for projection algorithms when the step size is sufficiently small and the problem function is strongly-f-monotone. Finally, we establish a similar result for a generalized steepest descent method for solving systems of equations (or unconstrained $VIPs$) when the Jacobian matrix is positive definite and the squared Jacobian matrix is positive semidefinite.

3.1 Averaging on a relaxation scheme

We consider a relaxation scheme that reduces the solution of the variational inequality problem to a succession of solutions of variational inequality problems with a simpler structure that can be solved by available efficient algorithms.
We consider a smooth function $g : K \times K \to R^n$ satisfying the condition that

$$g(x, x) = f(x) \quad \text{for all} \quad x \in K.$$

We also define $g(x, y)$ so that the matrix $g_x(x, y)$ is symmetric and positive definite.

A Relaxation Scheme

STEP 0:

Choose an arbitrary point $x_0 \in K$.

STEP $k+1$:

Find $x_{k+1} \in K$ satisfying the inequality

$$g(x_{k+1}, x_k)^t (x - x_{k+1}) \geq 0 \quad \forall x \in K. \tag{5}$$

The conditions imposed upon $g(.,.)$ imply that the variational inequality from step $k+1$ is equivalent to a strictly convex minimization problem with the objective function $F(x) = \int_0^x g(y, x_k) dy$. The original relaxation algorithms used by Ahn and Hogan [1] to compute equilibria in economic equilibrium problems used

$g_i(x_{k+1}, x_k) = f_i(x_k^1, ..., x_k^{i-1}, x_{k+1}^i, x_k^{i+1}, ..., x_k^n)$ for $i = 1, 2, ..., n$. This method is known as the PIES algorithm. Subsequently, Dafermos developed and analyzed a general relaxation scheme with the more general choice of g (as described above) in the context of both the traffic equilibrium problem [10] as well as the general variational inequality problem [11].

The papers [1], [10] and [11], and the references they cite, describe more details of these approaches. We wish to show that under appropriate assumptions, sequences of averages induced by the sequence $\{x_k\}_{k=1}^{\infty}$ converge to a solution of the original asymmetric VIP.

The following theorem summarizes the convergence results of Dafermos [10], [11] and of Ahn and Hogan [1].

THEOREM 3:

Let K be a convex, compact subset of R^n and consider the relaxation scheme. Let $T : K \rightarrow R^n$ be the map which carries a point in K to the solution of the relaxation scheme (5). Suppose that the algorithm function g satisfies the following conditions:

1. $g(x, x) = f(x)$.

2. The matrix $g_x(x, y)$ is positive definite and symmetric $\forall x, y \in K$.

3. If $\alpha = \inf_{x,y \in K}$ (min *eigenvalue* $g_x(x, y)$), then

$$\sup_{x,y \in K} ||g_y(x, y)|| \leq \lambda\alpha \text{ for some } 0 < \lambda < 1. \tag{6}$$

Then the sequence $\{T^k(x^0))\}$, with $x^0 \in K$, converges to the solution of the original variational inequality problem.

We now give an example that violates condition (6) and for which the relaxation algorithm does not converge.

Example:

Consider the variational inequality problem with problem function $f(x) = Mx$ with

$$M = \begin{bmatrix} b & -b \\ b & b \end{bmatrix}.$$

This matrix is asymmetric but positive definite. On the other hand, the matrix

$$g_x(x, y) = \begin{bmatrix} b & 0 \\ 0 & b \end{bmatrix}$$

is symmetric and positive definite. Moreover, $\alpha = \inf_{x,y \in K} (\min\ eigenvalue\ g_x(x, y)) = b > 0$, while

$$g_y(x, y) = \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}.$$

Then

$$||g_y(x, y)||^2 = \sup_{x \neq 0} \frac{x^t \begin{bmatrix} b^2 & 0 \\ 0 & b^2 \end{bmatrix} x}{x^t x} = b^2 = \alpha^2.$$

This problem satisfies the norm condition $||g_y(x, y)|| \leq \alpha$, but not the norm condition $||g_y(x, y)|| \leq \lambda\alpha$ for some $0 < \lambda < 1$.

Suppose we apply the relaxation scheme (5) to this problem function using as feasible a set the unit cube,

$K = \{x = (x_1, x_2) \in R^2 : -1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1\}$. The solution to this variational inequality problem is the point $x^* = (0, 0)$. Starting from the point $x^0 = (1, 1)$, the algorithm selects the points $x^1 = (1, -1)$, $x^2 = (-1, -1)$, $x^3 = (-1, 1)$ and $x^4 = x^1 = (1, 1)$. Therefore, the algorithm sequence cycles around the solution. Note, however, that the sequence of the averages induced by the relaxation algorithm converges to the solution point $x^* = (0, 0)$.

This example prompts us to relax the previous assumptions and establish results for the convergence of sequences of averages using the results of Section 2.

The main property we need to establish for the map T is that it is nonexpansive on

K.

Proposition 4:

Consider the algorithm function $g : K \times K \rightarrow R^n$ of the relaxation scheme (5). Suppose it satisfies the following conditions:

1. $g(x, x) = f(x)$.

2. The matrix $g_x(x, y)$ is positive definite and symmetric $\forall x, y \in K$.

3. If $\alpha = \inf_{x,y \in K}$ (min eigenvalue $g_x(x, y)$), then

$$\sup_{x,y \in K} ||g_y(x, y)|| \leq \alpha.$$

Then T is a nonexpansive map in K.

Proof:

To establish this result, we need to show that

$$||T(y_1) - T(y_2)|| \leq ||y_1 - y_2|| \quad \forall y_1, y_2 \in K.$$

Fix $y_1, y_2 \in K$ and set $T(y_1) = x_1$ and $T(y_2) = x_2$. Then the definition of T yields:

$$g(x_1, y_1)^t(x - x_1) \geq 0 \quad \forall x \in K, \tag{7}$$

$$g(x_2, y_2)^t(x - x_2) \geq 0 \quad \forall x \in K. \tag{8}$$

Setting $x = x_2$ in (7) and $x = x_1$ in (8) and adding the resulting inequalities, we obtain

$$[g(x_2, y_2) - g(x_1, y_1)]^t(x_1 - x_2) \geq 0. \tag{9}$$

By adding and subtracting $g(x_2, y_1)$, we can rewrite this expression as

$$[g(x_2, y_2) - g(x_2, y_1)]^t(x_1 - x_2) \geq [g(x_1, y_1) - g(x_2, y_1)]^t(x_1 - x_2). \tag{10}$$

Applying a mean value theorem on the righthand side of the inequality, we obtain

$$[g(x_2, y_2) - g(x_2, y_1)]^t(x_1 - x_2) \geq [x_1 - x_2]^t[g_x(x', y_1)][x_1 - x_2], \quad x' \in [x_1; x_2]. \tag{11}$$

Since the matrix $g_x(x,y)$ is positive definite and symmetric $\forall x, y \in K$ (by assumption) and $\alpha = \inf_{x,y\in K}$ (min *eigenvalue* $g_x(x,y)$),

$$[g(x_2,y_2) - g(x_2,y_1)]^t(x_1 - x_2) \geq \alpha||x_1 - x_2||^2. \tag{12}$$

Moreover, by applying a mean value theorem to the lefthand side of the inequality, we obtain:

$$[y_2 - y_1]^t[g_y(x_2,y')](x_1 - x_2) \geq \alpha||x_1 - x_2||^2, \quad y' \in [y_2; y_1]. \tag{13}$$

Furthermore, Cauchy's inequality and the operator norm inequality implies that

$$||y_1 - y_2||||g_y(x_2,y')||||x_1 - x_2|| \geq \alpha||x_1 - x_2||^2, \quad y' \in [y_2; y_1]. \tag{14}$$

Dividing both sides of this inequality by $||x_1 - x_2||$ gives

$$||y_1 - y_2||||g_y(x_2,y')|| \geq \alpha||x_1 - x_2||, \quad y' \in [y_2; y_1]. \tag{15}$$

Finally, the second assumption of this proposition, namely,

$$\sup_{x,y\in K} ||g_y(x,y)|| \leq \alpha,$$

implies that the map T is nonexpansive. This is true because this inequality implies that

$$\alpha||y_1 - y_2|| \geq \alpha||x_1 - x_2||. \tag{16}$$

Therefore, T is a nonexpansive map since,

$$||T(y_1) - T(y_2)|| \leq ||y_1 - y_2|| \quad \forall y_1, y_2 \in K.$$

Q.E.D.

Using this proposition, we now establish the convergence of sequences of averages induced by this relaxation algorithm.

Part (a) and part (b) in the following theorem use the finite dimensional versions of Theorems 1 and 2, respectively.

THEOREM 4:

Let K be a convex, closed subset of R^n (the feasible set of the original VIP) and let $T : K \longrightarrow R^n$ be the map that carries a point in K to the solution of the relaxation scheme (5).

Suppose the algorithm function g satisfies conditions of Proposition 4,

1. $g(x, x) = f(x)$.

2. The matrix $g_x(x, y)$ is positive definite and symmetric $\forall x, y \in K$.

3. If $\alpha = \inf_{x,y \in K}$ (min eigenvalue $g_x(x, y)$), then

$$\sup_{x,y \in K} ||g_y(x, y)|| \leq \alpha.$$

   (a) The sequence of averages

$$x_{k+1} = \frac{a_1 x_1 + a_2 T(x_1) + ... + a_k T(x_k)}{a_1 + ... + a_k} \quad x_1 \in K,$$

   $\sum_{k=1}^{\infty} \min(a(k), 1 - a(k)) = +\infty$, with $a(k) = \frac{a_k}{a_1 + ... + a_k}$, converges to a solution of the original asymmetric VIP.

   (b) Furthermore, the sequence of averages

$$S_k(y) = \frac{y + T(y) + ... + T^k(y)}{k + 1} \quad y \in K$$

   converges to a solution of the original asymmetric VIP.

Proof:

(a) Theorem 1 guarantees that the sequence of "Riesz" averages $\{x_k\}_k$ converges to an optimal solution of the VIP, since the map T is nonexpansive.

(b) The finite dimensional version of Theorem 2 guarantees that the sequence of averages

$$S_k(y) = \frac{y + T(y) + ... + T^k(y)}{k + 1} \quad y \in K,$$

converges to an optimal solution of the VIP, since the map T is nonexpansive. Q.E.D.

The next theorem gives the more general version of Theorem 4 which guarantees convergence of the averaging scheme induced by the relaxation algorithm.

THEOREM 5:

Let K be a convex, closed subset of R^n (the feasible set of the original VIP) and $T : K \longrightarrow R^n$ be the map that carries a point in K to the solution of the relaxation scheme (5).

Suppose the algorithm function g satisfies,

1. $g(x, x) = f(x)$.

2. The matrix $g_x(x, y)$ is positive definite and symmetric $\forall x, y \in K$.

3. For some positive definite matrix G, with symmetric part $S = \frac{G+G^t}{2}$, $g_x(x, y) - G$ is a positive semidefinite matrix for all $x, y \in K$ satisfying the condition

$$||S^{-1}[g(x, y_1) - g(x, y_2)]||_S \leq ||y_1 - y_2||_S$$

for all $x, y_1, y_2 \in K$.

(a) If $a(k) = \frac{a_k}{a_1+...+a_k}$, then the sequence of averages

$$x_{k+1} = \frac{a_1 x_1 + a_2 T(x_1) + ... + a_k T(x_k)}{a_1 + ... + a_k} \quad x_1 \in K,$$

with $\sum_{k=1}^{\infty} \min(a(k), 1 - a(k)) = +\infty$, converges to a solution of the original asymmetric VIP.

(b) Furthermore, the sequence of averages

$$S_k(y) = \frac{y + T(y) + ... + T^k(y)}{k+1} \quad y \in K,$$

converges to a solution of the original asymmetric VIP.

Proof: The proof is similar to that of Proposition 4 and Theorem 4.

Remarks:

1. Observe that in part (b) we used finite dimensional version of Baillon's Theorem. In the infinite dimensional version we need to assume that the feasible set K is compact. In the second part of Theorems 4 and 5 we do not assume that the feasible set K is compact. In this case, since the sequence $\{x_k\}$ is nonexpansive around any solution, if the problem has a solution x^*, then $||x_k|| \leq ||x_0 - x^*|| + ||x^*||$ and, therefore, the points x_k lie in a bounded set.

2. We do not need to assume the symmetry of the matrix $g_x(x, y)$. The analysis of Theorems 4 and 5 is valid provided that we replace the matrix $g_x(x, y)$ in Proposition 4 by its symmetric part.

3. If we set $G = \alpha I$, and $\alpha = \inf_{x,y \in K}$ (min eigenvalue $g_x(x, y)$), then the matrix $g_x(x, y) - \alpha I$ is a positive semidefinite matrix for all $x, y \in K$. In this case, the condition of Theorem 5, namely,

$$||S^{-1}[g(x, y_1) - g(x, y_2)]||_S \leq ||y_1 - y_2||_S,$$

for all $y_1, y_2, x \in K$, becomes

$$||g(x, y_1) - g(x, y_2)|| \leq \alpha ||y_1 - y_2||, \quad \text{for all } y_1, y_2, x \in K.$$

Observe that the weak norm condition of Theorem 4, $\sup_{x,y \in K} ||g_y(x, y)|| \leq \alpha$, also implies this condition.

For linearization schemes, which we study later (see [41] for a more detailed analysis), the norm condition of Theorem 5 becomes a more general version of the norm condition of the global convergence Theorem 2.9 of [41], that is, (i) for some positive definite matrix G, the matrix $A(y) - G$ is positive semidefinite for all $y \in K$, and (ii) for some constant $0 < b < 1$,

$$||S^{-1}[f(x) - f(y) - A(y)(x - y)]||_S \leq b||y - x||_S, \quad \forall y, x \in K.$$

In our case, b can be also 1.

4. In Section 2 we observed that we could require the nonexpansiveness of the map T only around the solutions x^*, provided that in the averaging scheme $\sum_{k_j} \min(a(k_j), 1 - a(k_j)) = +\infty$ for all subsequences k_j. In the case of the previous relaxation scheme, the norm condition of Theorem 5 around the solutions x^* becomes:

$$||S^{-1}[g(x^*, y) - f(x^*)]||_S \leq ||y - x^*||_S, \quad \forall y \in K, \tag{17}$$

and in the case of linearization schemes [41], it becomes:

$$||S^{-1}[f(x^*) - f(y) - A(y)(x^* - y)]||_S \leq ||y - x^*||_S, \quad \forall y \in K. \tag{18}$$

Therefore, we obtain global convergence results by requiring a norm condition (17) only around the solutions (or (18) in the case of linearization schemes). To establish local convergence results of linearization schemes, Pang and Chan [41] require (18) for a constant $0 < b < 1$. The initial iterate x_0 needs to satisfy the condition $||x_0 - x^*|| < \frac{1-b}{C}$, for some constant $C > 0$. Therefore, the closer b is to 1, the closer the initial point needs to be to a solution. Our proof (see Theorem 5) does not require this initial condition.

We will now try to provide some intuiton concerning the norm condition that we imposed in Theorem 4. The question we address is,

what is the relationship, if any, between the norm condition of Theorem 4 and conditions imposed upon the original problem function?

Pang and Chan have extensively studied linearization algorithms for solving $VIPs$. Linearization schemes also fit in the framework of the general iterative scheme developed by Dafermos [11], which in general works as in (5).

In the case of the general iterative scheme, we impose the following more general conditions on the scheme's function g:

1. $g(x, x) = f(x)$,

2. the Jacobian matrix of $g(x,y)$ with respect to the x component, $g_x(x,y)$, when evaluated at the point $y = x$, is a positive definite and symmetric matrix.

For linearization algorithms, $g(x,y) = f(y) + \frac{1}{\rho}A(y)(x-y)$ for some positive definite matrix $A(y)$ and constant $0 < \rho \leq 1$. In most cases, researchers have chosen ρ is equal to one.

In the context of these algorithms, the norm condition we need to impose (see Pang and Chan [41] for more details) is

$$||(g_x^{-1/2}(x,x))^t g_y(x,x) g_x^{-1/2}(x,x)|| < 1 \quad \forall x \in K, \tag{19}$$

which can also be rewritten (see Section 1) as follows:

$$||g_x^{-1}(x,x) g_y(x,x)||_{g_x(x,x)} < 1 \quad \forall x \in K.$$

In particular, for the linearization algorithms, since

$$g(x,y) = f(y) + \frac{1}{\rho}A(y)(x-y),$$

$g_x(x,x) = \frac{1}{\rho}A(x)$ is positive definite and symmetric ($A(x) = A(x)^t$), $g_y(x,x) = \nabla f(x) - \frac{1}{\rho}A(x)$, and so $\nabla f(x) = g_y(x,x) + g_x(x,x)$. The norm condition becomes

$$||(A(x)^{-1/2})^t[\rho\nabla f(x) - A(x)](A(x)^{-1/2})|| = ||I - \rho A^{-1/2}(x)\nabla f(x) A^{-1/2}(x)|| < 1.$$

Whenever $A(y) = \nabla f(y) + \nabla f(y)^t$ and $\rho = 1$, the norm condition becomes

$$||(\nabla f(x) + \nabla f(x)^t)^{-1/2}(\nabla f(x))^t(\nabla f(x) + \nabla f(x)^t)^{-1/2}|| =$$

$$= ||(\nabla f(x) + \nabla f(x)^t)^{-1}\nabla f(x)^t||_{\nabla f(x)+\nabla f(x)^t} < 1.$$

Notice that the norm condition used by Dafermos for the convergence of the general iterative scheme, namely,

$$||g_x^{-1/2}(x_1,y_1) g_y(x_2,y_2) g_x^{-1/2}(x_3,y_3)|| < 1 \quad \forall x_1,y_1,x_2,y_2,x_3,y_3 \in K,$$

includes the norm conditions of Pang and Chan as special cases. This condition is more difficult to verify, however, since it involves different points $x_1, y_1, x_2, y_2, x_3, y_3$.

The norm condition of Theorem 4,

$$\sup_{x,y \in K} ||g_y(x,y)|| \leq \alpha,$$

implies the norm condition (19) in a less than or equal form. This is true because

$$||g_x^{-1/2}(x,x)g_y(x,x)g_x^{-1/2}(x,x)|| \leq ||g_x^{-1/2}(x,x)||||g_y(x,x)||||g_x^{-1/2}(x,x)|| \leq$$

$$\leq \alpha^{-1/2}.\alpha.\alpha^{-1/2} = 1 \quad \forall x \in K$$

via the operator norm inequality and $\alpha = \inf_{x,y \in K}$ (min *eigenvalue* $g_x(x,y)) > 0$.

To provide some intuition concerning these norm conditions and the original problem function, we will now investigate their relationship to the strong-f-monotonicity condition. The next theorem shows that the differential form of strong-f-monotonicity of f implies the norm condition (19) in a more general form, a less than or equal form instead of a strictly inequality form. Furthermore, the theorem also demonstrates a partial converse of this statement. Namely, (19) implies a weaker form of the differential condition of strong-f-monotonicity.

Before analyzing the main theorem, we state and prove two useful lemmas.

LEMMA 3:

If A is a positive semidefinite matrix and G a positive definite, symmetric matrix, then $G^{-1/2}AG^{-1/2}$ is also a positive semidefinite matrix.

Proof:

If $x \in R^n$ and $y = G^{-1/2}x$, then $x^tG^{-1/2}AG^{-1/2}x = y^tAy \geq 0$ since A is a positive semidefinite matrix. Therefore, $x^tG^{-1/2}AG^{-1/2}x \geq 0$ for all $x \in R^n$ and so $G^{-1/2}AG^{-1/2}$ is a positive semidefinite matrix. Q.E.D.

LEMMA 4:

Suppose that the matrix

$$\nabla f(x)^t - a\nabla f(x)^t \nabla f(x), \forall x \in K$$

is positive semidefinite for some constant $a > 0$. Let G be a positive definite matrix, g be the minimum eigenvalue of G, and $a_1 \leq ag$. Then $(G^{-1/2}\nabla f(x)G^{-1/2})^t(I - a_1 G^{-1/2}\nabla f(x)G^{-1/2})$ is also positive semidefinite.

Proof:

Recall that for any vector $v \in R^n$, and for all $y \in R^n$,

$$y^t(G^{-1/2}\nabla f(x)G^{-1/2})^t(I - a_1 G^{-1/2}\nabla f(x)G^{-1/2})y \geq$$

(replacing $a_1 \leq ag$, and $z = G^{-1/2}y$, $v^t G^{-1} v \leq g v^t v$, we obtain)

$$\geq (G^{-1/2}y)^t[\nabla f(x)^t - ag\nabla f(x)^t G^{-1}\nabla f(x)](G^{-1/2}y) \geq$$

$$\geq z^t[\nabla f(x)^t - a\nabla f(x)^t \nabla f(x)]z \geq 0.$$

The last inequality follows from the assumption, and so the matrix $(G^{-1/2}\nabla f(x)G^{-1/2})^t(I - a_1 G^{-1/2}\nabla f(x)G^{-1/2})$ is positive semidefinite. Q.E.D.

LEMMA 5: (see also Proposition 6)

The matrix $B^t[I - (a/2)B]$ is positive semidefinite if and only if the operator norm $||I - aB|| \leq 1$. Moreover, if both conditions are satisfied for any value a^* of a, then they are satisfied for all values $a \leq a^*$.

Proof:

Recall that

$$||I - aB|| = \sup_{y \neq 0} \frac{||(I - aB)y||^2}{||y||^2} \leq 1.$$

Therefore,

$$\begin{aligned}
&||I - aB|| \leq 1 \\
\Leftrightarrow &\sup_{y \neq 0} \frac{y^t[I - (aB^t + aB) + (aB)^t(aB)]y}{y^t y} \leq 1 \\
\Leftrightarrow &\; y^t[I - (aB^t + aB) + (aB)^t(aB)]y \leq y^t y \quad \forall y \in R^n \\
\Leftrightarrow &\; 2ay^t By \geq a^2 y^t B^t By \quad \forall y \in R^n \\
\Leftrightarrow &\; y^t By \geq (a/2)y^t B^t By \quad \forall y \in R^n \qquad (20) \\
\Leftrightarrow &\; y^t B^t[I - (a/2)B]y \geq 0 \quad \forall y \in R^n
\end{aligned}$$

These relationships show that $||I - aB|| \leq 1$ if and only if the matrix $B^t[I - (a/2)B]$ is positive semidefinite. Moreover, (20) implies that if both conditions are valid for any value a^* of a, then they are valid for all values $a \leq a^*$. Q.E.D.

This Lemma also is valid in another form: $B^t[I - (a/2)B]$ positive definite if and only if $||I - aB|| < 1$.

We are now ready to prove a theorem relating the norm codition to the differential form of strong-f-monotonicity.

THEOREM 6:

Consider the general iterative scheme and assume that $g_x(x, x)$ is a positive definite and symmetric matrix. Then the following results are valid.

1. If the differential form of the strong-f-monotonicity condition holds for a constant $a > 0$ and if $1 \leq 2g_{min}a$ for $g_{min} = \inf_{x \in K}$ [min *eigenvalue* $g_x(x, x)$], then the norm condition holds in a less than or equal to form (that is, expression (19) with $\leq$ instead of $<$).

2. Conversely, if the norm condition (19) holds in a less than or equal to form, then for some constant $0 < a \leq \frac{1}{2g_{max}}$, where $g_{max} = \sup_{x \in K}$ [max *eigenvalue* $g_x(x, x)$], the matrix $\nabla f(x)^t - a\nabla f(x)^t \nabla f(x)$ is positive semidefinite for all $x \in K$.

Proof:

1. We want to show that the following norm condition holds:

$$||g_x^{-1/2}(x, x) g_y(x, x) g_x^{-1/2}(x, x)|| \leq 1 \quad \forall x \in K.$$

Since

$$g_y(x, x) = \nabla f(x) - g_x(x, x),$$

if we let $G = g_x(x, x)$, the norm condition becomes:

$||G^{-1/2}[\nabla f(x) - G]G^{-1/2}|| = ||I - G^{-1/2}\nabla f(x) G^{-1/2}|| \leq 1.$

By assumption, G is a positive definite and symmetric matrix. Let

$$g_{min} = \inf_{x \in K} [\text{min } \textit{eigenvalue } G],$$

which is positive since K is a compact set. Also, let $B = G^{-1/2}\nabla f(x)G^{-1/2}$. Lemma 4 shows that if $a_1 = ag_{min}$, the matrix

$$B^t[I - a_1 B] = (G^{-1/2}\nabla f(x)G^{-1/2})^t(I - a_1 G^{-1/2}\nabla f(x)G^{-1/2})$$

is positive semidefinite. Lemma 5 implies that if

$$0 < 1 \leq 2a_1 = 2ag_{min}, \quad ||I - B|| \leq 1.$$

Making the replacement $B = G^{-1/2}\nabla f(x)G^{-1/2}$, we see that for $0 < 1 \leq 2a_1 = 2ag_{min}$,

$$||G^{-1/2}g_y(x,x)G^{-1/2}|| = ||G^{-1/2}[\nabla f(x) - G]G^{-1/2}|| =$$

$$= ||I - G^{-1/2}\nabla f(x)G^{-1/2}|| \leq 1, \quad \forall x \in K.$$

Therefore, for $G = g_x(x,x)$,

$$||g_x^{-1/2}(x,x)g_y(x,x)g_x^{-1/2}(x,x)|| \leq 1, \quad \forall x \in K.$$

2. In the second part of the theorem we want to prove that if the norm condition

$$||g_x^{-1/2}(x,x)g_y(x,x)g_x^{-1/2}(x,x)|| \leq 1, \quad \forall x \in K,$$

holds, then the matrix

$$\nabla f(x)^t - a\nabla f(x)^t\nabla f(x),$$

is positive semidefinite for some $a > 0$ and $\forall x \in K$.

Let $G = g_x(x,x)$. Since

$$||g_x^{-1/2}(x,x)g_y(x,x)g_x^{-1/2}(x,x)|| = ||G^{-1/2}[\nabla f(x) - G]G^{-1/2}|| =$$

$$= ||I - G^{-1/2}\nabla f(x)G^{-1/2}|| \leq 1, \quad \forall x \in K,$$

setting, as before, $B = G^{-1/2}\nabla f(x)G^{-1/2}$, we see from Lemma 5 that if

$$||I - B|| \leq 1,$$

for any value $a_1 \leq 1/2$, then the matrix $B^t[I - a_1 B]$ is positive semidefinite. Let

$$g_{max} = \sup_{x \in K}[\max \ eigenvalue \ G].$$

Then if $1 \geq 2a_1 \geq 2ag_{max}$,

$$y^t B^t y \geq a_1 y^t B^t B y \geq a y^t B^t G B y \quad \forall y \in R^n.$$

Making the replacement $B = G^{-1/2} \nabla f(x) G^{-1/2}$, we obtain

$$y^t [G^{-1/2} \nabla f(x)^t (I - a \nabla f(x)) G^{-1/2}] y \geq 0.$$

Finally, setting $z = G^{-1/2} y$, we see that $z^t [\nabla f(x)^t (I - a \nabla f(x))] z \geq 0$.
These results show that for any $a \leq \frac{1}{2g_{max}}$, the matrix

$$\nabla f(x)^t (I - a \nabla f(x)) \quad \forall x \in K$$

is positive semidefinite. Q.E.D.

Remark:

The differential condition of strong-f-monotonicity implies that the norm condition holds in a less than or equal form. [41] requires a strict inequality form of the norm condition. This happens when the differential form of strong-f-monotonicity holds in some form of a strict inequality, i.e.,

$$[\nabla f(x)^t (I - a \nabla f(y))]$$

is positive semidefinite and the matrix $\nabla f(x)$ is nonsingular. The norm condition (19) then holds as a strict inequality. Therefore, (19) implies that the original problem function f is strictly monotone. Finally, we would like to note that when the norm condition (19) becomes:
$||g_x^{-1/2}(x,x) g_y(x,x) g_x^{-1/2}(x,x)|| \leq \lambda < 1 \quad \forall x \in K$, then this condition becomes equivalent to the differential form of strong monotonicity on the problem function f.

The following Proposition formalizes this result.

Proposition 5:

Consider the general iterative scheme in which $g(x,x) = f(x)$ and $g_x(x,x)$ is a positive definite and symmetric matrix. Then the following results are valid.

1. If the differential form of strong-f-monotonicity condition holds as a strict inequality, i.e., the matrix

$$\nabla f(x)^t - a\nabla f(x)^t \nabla f(y) \quad \forall x, y \in K,$$

is positive semidefinite for some constant $a > 0$, and the matrix $\nabla f(x)$ is nonsingular, then the norm condition (19) holds as a strict inequality.

2. If the differential form of the strong monotonicity condition holds, then for some $0 < \lambda < 1$ the following norm condition holds,

$$||g_x^{-1/2}(x,x)g_y(x,x)g_x^{-1/2}(x,x)|| \leq \lambda \quad \forall x \in K.$$

3. If for some $0 < \lambda < 1$, the norm condition (19) holds, then for some constant $a > 0$, the matrix

$$\nabla f(x)^t - a\nabla f(x)^t \nabla f(x)$$

is positive definite $\forall x \in K$.

4. If the norm condition

$$||g_x^{-1/2}(x,x)g_y(x,x)g_x^{-1/2}(x,x)|| \leq \lambda \quad \forall x \in K,$$

holds, then the differential form of the strong monotonicity condition holds.

The proof of this Proposition is similar to that of Theorem 6.

This discussion shows that various forms of the norm condition (19) have "equivalent", in some sense, formulations as monotonicity conditions (in their differential forms).

For linearization algorithms, the underlined conditions behind their convergence to the optimal solution of the variational inequality problem, are the differential form of strong-f-monotonicity and the assumption that $\nabla f(x)$ is positive definite.

3.2 Averaging for the projection algorithm

Since fixed points of the map $T(x) = Pr_K^G(x - \rho G^{-1} f(x))$ are identical to solutions to the variational inequality problem $VI(f, K)$, and since the projection operator $Pr_K^G(y)$ is nonexpansive with respect to the G-norm, the mapping $T(x)$ will be nonexpansive whenever the map $T'(x) = x - \rho G^{-1} f(x)$ is. As noted by Magnanti and Perakis [29], if $G = I$, the identify matrix, then $T'(x)$ is a nonexpansive mapping, (which they state as Lipschitz continuity over K for a Lipschitz constant of 1) if and only if the map f is strongly-f-monotone with respect to the monotonicity constant $a = \rho/2$. The following result generalizes this observation.

Proposition 6:

Consider the variational inequality problem $VI(f, K)$. Let G be a positive definite, symmetric matrix and let g_{min} and g_{max} be its smallest and largest eigenvalues. The map $T'(x) = x - \rho G^{-1} f(x)$ is nonexpansive on the feasible set K, with respect to the G-norm, if the map f is strongly-f-monotone with respect to the constant $a \geq \frac{\rho}{2g_{min}}$. Conversely, if the map T' is nonexpansive on the feasible set K, with respect to the G-norm then the map f is strongly-f-monotone with respect to the constant $a = \frac{\rho}{2g_{max}}$. In the special case $G = I$, $g_{min} = g_{max} = 1$ and, therefore, $a = \rho/2$, as stated in [29].

Proof: First observe that

$$||T'(x) - T'(y)||_G^2 = ||x - \rho G^{-1} f(x) - y + \rho G^{-1} f(y)||_G^2 =$$

$$||x - y||_G^2 + \rho^2 (f(x) - f(y))^t G^{-1} (f(x) - f(y)) - 2\rho (f(x) - f(y))^t (x - y). \quad (21)$$

" $\Rightarrow$" Let g_{min} be the smallest eigenvalue of the matrix G. Then, (21) implies that

$$||T'(x) - T'(y)||_G^2 \leq ||x - y||_G^2 + \frac{\rho^2}{g_{min}} ||f(x) - f(y)||^2 - 2\rho (f(x) - f(y))^t (x - y).$$

Therefore, if f is a strongly-f-monotone map with respect to the constant $a \geq \frac{\rho}{2g_{min}}$, then the map T' is nonexpansive.

" $\Leftarrow$" Conversely, if the map T' is nonexpansive, then using (21) and a similar

argument, we conclude that f is a strongly-f-monotone map with respect to the constant $a = \frac{\rho}{2g_{max}}$ and the largest eigenvalue g_{max} of the matrix G. Q.E.D.

Corollary 3:

Let g_{min} be the smallest eigenvalue of the matrix G. If the map f is Lipschitz continuous with constant $L > 0$ and strongly monotone with constant $a \geq \frac{\rho L^2}{2g_{min}}$, then the map $T'(x) = x - \rho G^{-1} f(x)$ is a contraction on the feasible set K, with respect to the G-norm, with constant $0 < b^2 = 1 - \frac{\rho}{g_{max}}(2a - \frac{\rho L^2}{g_{min}}) < 1$. Conversely, if the map T' is contraction on the feasible set K, with respect to the G-norm with constant $0 < b < 1$, then the map f is strongly monotone with constant $a = \frac{(1-b^2)g_{min}}{2\rho}$.

The proof of this corollary is similar to that of Proposition 6.

We now establish the convergence of sequences of averages induced by the projection algorithm. To establish these results, we first recall the projection algorithm.

The Projection Algorithm

Fix a positive definite and symmetric matrix G and a positive scalar ρ, whose value we will select below.

STEP 0:

Start with some $x_0 \in K$.

STEP $k+1$:

Compute $x_{k+1} \in K$ by solving the variational inequality VI_k:

$$[\rho f(x_k) + G(x_{k+1} - x_k)]^t (x - x_{k+1}) \geq 0, \quad \forall x \in K. \tag{22}$$

If we let $Pr_K^G(y)$ denote the projection of the vector y onto the feasible set K, with respect to the $||.||_G$ norm, we can view this step as the following projection operation:

$$x_{k+1} = Pr_K^G(x_k - \rho G^{-1} f(x^k)).$$

Note that this algorithm is a special case of the general iterative scheme [11] and the linearization algorithms [41], with $g(x, y) = \rho f(y) + A(y)(x - y)$ and $A(y) = G$.

In this section we establish the convergence of sequences of averages induced by the projection algorithm.

For the subsequent analysis, let us define $T_\rho : K \to R^n$ as the map that carries $x_k \in K$ into the minimizer over K of the function $F_k(.)$. Therefore, $x_{k+1} = T_\rho(x_k)$. The following lemma describes the relevance of the map T_ρ and shows the connection between the variational inequality problem and a fixed point problem.

LEMMA 6: (see Dafermos [9])

Every fixed point of the map T_ρ is a solution of the original asymmetric variational inequality problem.

Theorems 1 and 2 require the nonexpansiveness of map T_ρ, which we establish in the following lemma.

LEMMA 7:

Let g_{min} be the minimum eigenvalue of the positive definite, symmetric matrix G, and $b = \frac{1}{g_{min}}$. If $0 < \rho \leq \frac{2a}{b}$ and f is a strongly-f-monotone map (with respect to the constant a), then the map T_ρ is a nonexpansive map on the feasible set K with respect to the norm $||x||_G = (x^t G x)^{1/2}$. That is,

$$||T_\rho(y_1) - T_\rho(y_2)||_G \leq ||y_1 - y_2||_G \quad \forall y_1, y_2 \in K.$$

Proof: Let $y_1, y_2 \in K$ and set $x_1 = T_\rho(y_1)$ and $x_2 = T_\rho(y_2)$. The definition of map T_ρ shows that

$$(Gx_1 + h_1)^t(x - x_1) = (Gx_1 + \rho f(y_1) - Gy_1)^t(x - x_1) \geq 0 \quad \forall x \in K, \tag{23}$$

$$(Gx_2 + h_2)^t(x - x_2) = (Gx_2 + \rho f(y_2) - Gy_2)^t(x - x_2) \geq 0 \quad \forall x \in K. \tag{24}$$

Setting $x = x_2$ in (23) and $x = x_1$ in (24) and adding the two inequalities, we see that

$$||x_1 - x_2||_G^2 \leq \{y_1 - y_2 - \rho G^{-1}[f(y_1) - f(y_2)]\}^t G(x_1 - x_2).$$

Applying Cauchy's inequality, we find that

$$||x_1 - x_2||_G^2 \leq ||y_1 - y_2 - \rho G^{-1}[f(y_1) - f(y_2)]||_G ||x_1 - x_2||_G.$$

Dividing through by $||x_1 - x_2||_G$, squaring, and expanding the righthand side, we obtain

$$||x_1 - x_2||_G^2 \leq ||T'(y_1) - T'(y_2)||_G^2,$$

where $T'(x) = x - \rho G^{-1} f(x)$ is the map we define in Proposition 6. Therefore, as we show in Proposition 6, the strong-f-monotonicity of map f and the symmetry and positive definiteness of matrix G, together with this result, imply that if $0 < \rho \leq \frac{2a}{b}$ then

$$||T_\rho(y_1) - T_\rho(y_2)||_G = ||x_1 - x_2||_G^2 \leq ||y_1 - y_2||_G^2.$$

Q.E.D.

THEOREM 7:

Let K be a convex, closed subset of R^n (the feasible set of the VIP and $T_\rho : K \rightarrow R^n$ be the map that carries $x \in K$ into the solution of (22). Also, let λ be the minimum eigenvalue of the positive definite, symmetric matrix G and let $b = \frac{1}{g_{min}}$.

1. Let $a(k) = \frac{a_k}{a_1 + ... + a_k}$ and $a_k > 0$ be given constants. Assume in the projection algorithm that $0 < \rho \leq \frac{2a}{b}$. Then, if f is a strongly-f-monotone map, the sequence of averages

$$x_{k+1} = \frac{a_1 x_1 + a_2 T_\rho(x_1) + ... + a_k T_\rho(x_k)}{a_1 + ... + a_k}, \quad x_1 \in K,$$

where $\sum_{k=1}^{\infty} \min(a(k), 1 - a(k)) = +\infty$, converges to a solution of the original asymmetric variational inequality problem.

2. If $0 < \rho \leq \frac{2a}{b}$, then the strong-f-monotonicity of map f implies that the sequence of averages

$$S_\rho^k(y) = \frac{y + T_\rho(y) + ... + T_\rho^{k-1}(y)}{k}, \quad y \in K$$

converges to a solution of the original asymmetric variational inequality problem.

Proof: From Lemma 7, if $0 < \rho \leq \frac{2a}{b}$ in the projection algorithm and if f is strongly-f-monotone, then the map T_ρ is nonexpansive relative to the $||.||_G$ norm.

1. The finite dimensional version of Theorem 2 guarantees that

$$S_\rho^k(y) = \frac{y + T_\rho(y) + \ldots + T_\rho^{k-1}(y)}{k}, \; y \in K$$

converges to the fixed point of the map T_ρ. Lemma 6 shows that every fixed point of the map T_ρ is a solution of the original asymmetric variational inequality problem.

2. If we assume that the feasible set is closed and convex, by applying Theorem 1 we can establish convergence for sequences of "Riesz" type averages.

Q.E.D.

Remarks:

1. If we choose $0 < \rho < \frac{2a}{b}$ and the function f is one-to-one, then T_ρ is not only a nonexpansive map, but also a contractive map. In this case, the original sequence induced by the projection algorithm converges to the the solution of the VIP, which is unique (since f is strictly monotone). The convergence of the original sequence follows from Banach's fixed point theorem (see also [9], [11]).

2. Observe that the projection algorithm, described in this subection, is a special case of the relaxation scheme (5) for the choice $g(x, y) = \rho f(y) + G(x - y)$. Then $g_x(x, y) = G$ which is positive definite and symmetric. Furthermore, choosing $S = G$ in Theorem 5, we observe that the norm condition of Theorem 5 is valid if the problem function f is strongly-f-monotone and ρ is chosen as in Theorem 7.

If we assume that the feasible set K is a closed and convex set and that $\rho < \frac{2a}{b}$, then similar results hold for the sequence (not the averages) induced by the projection algorithm.

THEOREM 8: (see Gabay [14])

If the variational inequality problem has at least one solution x^*, the feasible set K

is a closed, convex subset of R^n, the problem function f is strongly-f-monotone, and $0 < \rho < \frac{2a}{b}$ in the projection algorithm, then the limit of the algorithm sequence solves the variational inequality problem.

Remarks:

(a) Theorem 7 is an application of Theorem 1, developed in Section 2 and the ergodic theorem of J.B. Baillon [3]. Using these theorem, establishes the convergence of the sequences of averages induced by the algorithmic map T. On the other hand, Theorem 8 is based on a convergence theorem established in [34] by Z. Opial for solving fixed point problems of nonexpansive maps. This theorem, which does not consider averages, establishes the convergence of the original sequence under certain assumptions.

Opial's Theorem (Opial [34])

Let T be a map, $T : K \rightarrow K$, defined on a closed, and convex subset K of R^n. If T is a *nonexpansive* map on K relative to the $||.||_G$ norm, and for every point $y \in K$, T is *asymptotically regular*, that is, $||T^{k+1}(y) - T^k(y)||_2 \longrightarrow_{k\rightarrow\infty} 0$, then the map $T^k(y)$ converges to a fixed point of map T.

This theorem not only requires nonexpansiveness of map T, but also requires the additional property of asymptotic regularity. Nevertheless, it does not require the boundedness of the feasible set K.

Finally, we observe that if the map T satisfies the condition of firmly nonexpansiveness (used, for example, by Lions and Mercier [25], Rockafellar [45] and Bertsekas and Eckstein [12]) then it satisfies the asymptotic regularity condition. The convergence of the original sequence then follows from Opial's lemma.

Definition 2 *: A mapping $T : K \rightarrow K$ is firmly nonexpansive (or pseudocontractive) over the set K if*

$$||T(x) - T(y)||^2 \leq ||x - y||^2 - ||[x - T(x)] - [y - T(y)]||^2 \quad \forall x, y \in K$$

Expanding $||[x-T(x)]-[y-T(y)]||^2$ as $||x-y||^2+||T(x)-T(y)||^2-2[T(x)-T(y)]^t[x-y]$ and rearranging shows that T is strongly-f-monotone with coefficient $a = 1$, that is,

$[T(x) - T(y)]^t[x - y] \geq ||T(x) - T(y)||^2$.

Corollary 3:

If $T : K \longrightarrow K$ is a firmly nonexpansive map, then the original sequence induced the map T, that is $\{z_k = T(z_{k-1})\}_k$ converges to the fixed point of T.

Proof: Since firmly nonexpansive maps are nonexpansive, $||z_k - x^*||_G \leq ||z_{k-1} - x^*||_G$ for any fixed point x^* of the map T. This result together with the firmly nonexpansiveness condition of T implies that

$$||T(z_k) - z_k||_G^2 = ||z_k - x^*||_G^2 + ||T(z_k) - x^*||_G^2 - 2(T(z_k) - T(x^*))^t(z_k - x^*) \leq$$

$$||z_k - x^*||_G^2 - ||T(z_k) - x^*||_G^2 \longrightarrow_{k\to\infty} 0.$$

This property is the asymptotic regularity property of the map T. Therefore, Opial's lemma [34] implies the convergence of the sequence z_k. Q.E.D.

(b) It is important to choose $0 < \rho < \frac{2a}{b}$ in the projection algorithm to ensure that the sequence itself converges to a solution. The example of the previous section $f(x) = (bx_1 - bx_2, bx_2 + bx_1)$ illustrates this point. In this case, the strong-f-monotonicity constant is $a = \frac{1}{2}$, $\frac{2a}{b} = 1 = \rho$, and so Theorem 8 does not apply. As we have already shown when we introduced this example, the sequence of averages considered in Theorem 7 converges to a solution while the sequence itself cycles. If $G = I$, the identity matrix, and constant $\rho = 1$, then the projection algorithm is equivalent to the relaxation scheme that we considered in this example.

3.3 Averaging for the generalized steepest descent method

In this subsection we establish the convergence of sequences of averages induced by the generalized steepest descent method [17] applied to the unconstrained VIP with the underlying set $K = R^n$; this problem is equivalent to the system of equations $f(x^*) = 0$.

The Generalized Steepest Descent Method

STEP 0:

Start with some $x_0 \in R^n$.

STEP $k+1$:

Direction Choice:

Compute $-f(x_k)$. If $f(x_k) = 0$, stop; $x_k = x^*$. Otherwise continue.

One-Dimensional VIP:

Find $x_{k+1} \in [x_k; -f(x_k)]$ satisfying

$$f(x_{k+1})^t(x - x_{k+1}) \geq 0, \quad \forall x \in [x_k; -f(x_k)]. \tag{25}$$

In this description, $[x; d]$ denotes the ray emanating from x in the direction d: i.e., $[x; d] = \{y : \ y = x + ld, \ l \geq 0\}$.

The following theorem summarizes the convergence results of Hammond and Magnanti [17].

THEOREM 9:

Let M be a positive definte matrix and $f(x) = Mx - b$. Then the sequence of iterates induced by the generalized steepest descent method contracts (with respect to the $||.||_S$ norm induced by the matrix $S = \frac{M+M^t}{2}$) to the solution x^* of the unconstrained VIP if and only if the matrix M^2 is positive definite.

We establish the convergence of a sequence of averages induced by this method to the solution.

THEOREM 10:

Let $a(k) = \frac{a_k}{a_1+...+a_k}$, $a_k > 0$ be given constants, let M be a positive definte matrix, and assume $f(x) = Mx - b$. Consider the unconstrained VIP and let $T : R^n \rightarrow R^n$ be the map that carries x into the solution of (25). Suppose $\sum_{\{k_j\}_{\in N}} \min(a(k), 1 - a(k)) = +\infty$, for any subsequence $\{k_j\}_{\in N}$ of k, then if M^2 is a positive semidefinite matrix, the sequence of averages

$$x_{k+1} = \frac{a_1 x_1 + a_2 T(x_1) + ... + a_k T(x_k)}{a_1 + ... + a_k},$$

converges to a solution of the asymmetric unconstrained variational inequality problem.

Proof: We first observe that $T(x) = x - a(x)f(x)$, with $a(x) = \frac{||f(x)||^2}{||f(x)||_S^2}$ and $S = \frac{M+M^t}{2}$. The fixed points of the map T are also the solutions of the unconstrained VIP. When M^2 is positive semidefinite matrix

$$||T(x) - T(x^*)||_S^2 = ||x - x^*||_S^2 - a(x)(x - x^*)^t M^2 (x - x^*) \leq ||x - x^*||_S^2.$$

Therefore, T is a nonexpansive map around the solutions x^*. Remark 4 of Section 2 implies the convergence of the averaging scheme

$$x_{k+1} = \frac{a_1 x_1 + a_2 T(x_1) + ... + a_k T(x_k)}{a_1 + ... + a_k},$$

whenever $a(k) = \frac{a_k}{a_1 + ... + a_k}$, $a_k > 0$, and $\sum_{\{k_j\}_{j \in N}} \min(a(k), 1 - a(k)) = +\infty$ for any subsequence $\{k_j\}_{j \in N}$ of k. Q.E.D.

Remark:

As an example of such an $a(k)$, we can select $a_k = a^k$, with $a > 1$; then $a(k) = \frac{a^k(1-a)}{a - a^k}$.

Observe that Baillon's Theorem does not apply in this case since it requires nonexpansiveness of map T around every point and not just around the solutions.

Example:

Consider the unconstrained VIP with $f(x) = Mx$ and

$$M = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

This matrix is asymmetric but positive definite. The matrix

$$M^2 = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$$

is positive semidefinite. This example does not satisfy the assumptions of Theorem 9. In fact, if we initiate the steepest descent algorithm at the point $x_0 = [1, 1]$,

the algorithm cycles between the four points $[1,1]$, $[1,-1]$, $[-1,-1]$ and $[-1,1]$. Nevertheless, the assumptions of Theorem 10 hold and so, if $a_k = a^k$, with $a > 1$, the averaging scheme we proposed in Section 2 converges to the solution, which is in this case the point $x^* = [0,0]$.

4 Applications in transportation networks

In this section we apply the results from the previous sections to transportation networks. We first briefly outline the traffic equilibrium problem.

Consider a network G with links denoted by i, j,..., paths by p, q,... and origin-destination (O-D) pairs of nodes by w, z,... . A fixed travel demand, denoted d_w, is prescribed for every O-D pair w of the transportation network. Let F_p denote the nonnegative flow on path p. We group together all the path flows into a vector $F \in R^N$ (N is the total number of paths in the network). The travel demand d_w associated with the typical O-D pair w is distributed among the paths of the network that connect w. Thus,

$$d_w = \sum_{p \text{ joining } w} F_p, \quad \text{for all } O-D \ \ pair \ w, \tag{26}$$

or, in vector form, $d = BF$, where B is a $W \times N$ O-D pair/path incidence matrix whose (w,p) entry is 1 if path p connects O-D pair w and is 0 otherwise. The path flow F induces a load vector f with components f_i defined on every link i via

$$f_i = \sum_{p \ \text{passing through} \ i} F_p, \tag{27}$$

or, in vector form, $f = DF$, where D is a $n \times N$ link/path incidence matrix whose (i,p) entry is 1 if link i is contained in path p and is 0 otherwise. Let n be the total number of links in the network.

A load pattern f is feasible if some nonnegative path flow F, that is,

$$F_p \geq 0 \quad \text{for all paths} \ \ p, \tag{28}$$

induces the link flow f through (27) and is connected to the demand vector d through (26). It is easy to see that the set of feasible load patterns f is a compact, convex subset K of R^n.

Our goal is to determine the user optimizing traffic pattern with the equilibrium property that once established, no user can decrease his/her travel cost by making a unilateral decision to change his/her route. Therefore, in a user-optimizing network, the user's criterion for selecting a travel path is personal travel cost. We assume that each user on link i of the network has a travel cost c_i that depends, in an a priori specified fashion, on the load pattern f, and that the link costs vector $c = c(f)$ is a continuously differentiable function, $c : K \rightarrow R^n$. Finally, we let $C_p = C_p(F)$ denote the cost function on path p. The link and path cost functions are related as follows:

$$C_p(F) = \sum_{i \in path\ p} c_i(f), \quad \forall \text{ paths } p. \tag{29}$$

Mathematically, a flow pattern is a user equilibrium flow pattern if

$$\forall w \ (O\text{-}D\ pair),\ \forall p \text{ connecting } w : \ C_p(f) = v_w \text{ if } F_p > 0 \text{ and } C_p(f) \geq v_w \text{ if } F_p = 0.$$

The user equilibrium property can also be cast as the following variational inequality:

$$f^* \in K \text{ is user optimized if and only if } c(f^*)^t(f - f^*) \geq 0, \quad \forall f \in K. \tag{30}$$

Several papers [9], [10], [26], [47], [16] and the references they cite elaborate in some detail on this model and its extensions.

The analysis in the previous sections applies to the traffic equilibrium problem, with the travel cost function c as the VIP function and with the link flow pattern f as the problem variable. The strong-f-monotonicity condition becomes

$$[c(f^1) - c(f^2)]^t[f^1 - f^2] \geq a||c(f^1) - c(f^2)||_2^2 \quad \forall f^1, f^2 \in K,$$

for some positive constant a. As indicated in [27], we can verify this condition by checking whether the matrix

$$\nabla c(f)^t - a\nabla c(f)^t \nabla c(f') \quad \forall f, f' \in K$$

is positive semidefinite for some $a > 0$. Theorems 4, 5 and 7 guarantee that the sequence of averages of the "Riesz" type induced by relaxation and projection algorithms converges to an equilibrium solution f^* of the user optimizing network. Furthermore, since the feasible set K in the traffic equilibrium example is always bounded for any a priori fixed demand d, the second parts of Theorems 4, 5 and 7 also establish that the limit of the sequences of averages induced by the projection algorithm as in [3] is a user optimizing load pattern whenever our choice of $\rho \leq \frac{2a}{b}$. Finally, Theorem 8 establishes convergence for the sequence itself when $0 < \rho < \frac{2a}{b}$.

Next, we study some traffic equilibrium examples that illustrate the importance of the strong-f-monotonicity condition and allow us to apply the results in the previous sections.

Examples:

1. The simplest case arises when the travel cost function $c_i = c_i(f)$ on every link i depends solely, and linearly, upon the flow f_i on that link i:

$$c_i = c_i(f_i) = g_i f_i + h_i.$$

In this expression, g_i and h_i are nonnegative constants; g_i denotes the congestion coefficient for link i. Then $c(f) = \begin{bmatrix} g_1 & 0 & \dots & 0 \\ 0 & g_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & g_n \end{bmatrix} \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} + \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}$.

Since $\nabla c = \begin{bmatrix} g_1 & 0 & \dots & 0 \\ 0 & g_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & g_n \end{bmatrix}$, the matrix $\nabla c(f)^t - a\nabla c(f)^t \nabla c(f')$ becomes

$$\nabla c^t(I - a\nabla c) = \begin{bmatrix} g_1 - ag_1^2 & 0 & \dots & 0 \\ 0 & g_2 - ag_2^2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & g_n - ag_n^2 \end{bmatrix}.$$

This matrix is positive semidefinite if $g_i - ag_i^2 \geq 0$ for $i = 1, 2, ..., n$.
This, in turn, is true if the congestion coefficients $g_i \geq 0$ for $i = 1, 2, ..., n$ and $a \leq \frac{1}{\max_{1 \leq i \leq n} g_i}$.
If each $g_i > 0$ and $a < \frac{1}{\max_{1 \leq i \leq n} g_i}$, he matrix is positive definite, and so the function c is strongly monotone. It is positive semidefinite, and so is strongly-f-monotone even if some $g_i = 0$. Our analysis still applies even though some or all the $g_i's$ are zero. This example shows that strong-f-monotonicity might permit some links of the network to be uncongested. This might very well be the case in large scale networks.

2. We conclude this set of examples by considering a transportation network with multiple equilibria, as specified by a network (see Figure 1) consisting of one O-D pair $w = (x, y)$ and three links connecting this O-D pair. The travel demand is $d_w = 20$. The travel costs on the links are

$$c_1(f) = f_1 + f_2 + 5,$$

$$c_2(f) = f_1 + f_2 + 5,$$

$$c_3(f) = 30.$$

The user equilibrium solution is not unique. In fact, the problem has infinitely many user optimized solutions. Any point satisfying $f_1 + f_2 = 20$ and $f_3 = 0$

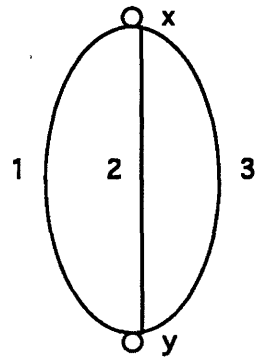

Figure 1: The traffic equilibrium problem

is a solution to the user optimized problem, since $c_1 = c_2 = 25 \leq c_3 = 30$.

The matrix $\nabla c = M$ is

$$M = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

This matrix is not positive definite. Nevertheless, the matrix

$$M^t - aM^tM = \begin{bmatrix} 1-2a & 1-2a & 0 \\ 1-2a & 1-2a & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

is positive semidefinite for any $a \leq 1/2$. So the travel cost function c in this case is strongly-f-monotone, but not strongly monotone.

We conclude this section by showing that if the link cost function is strongly-f-monotone, then so is the path cost function. In establishing this result, we use the following elementary lemma.

LEMMA 8:

Any set of n real numbers $x_i \in R$ for $i = 1, 2, ..., n$ satisfy the following inequality:

$$[\sum_{i=1}^{n}(x_i)]^2 \leq n\sum_{i=1}^{n}(x_i)^2. \tag{31}$$

This result is easy to establish by induction.

Proposition 7:

Let n be the total number of links in the network, and N be the total number of paths. If the link cost function $c = c(f)$ is strongly-f-monotone with respect to the constant a, then the path cost function $C = C(F)$ is also strongly-f-monotone with respect to the constant $a' = \frac{a}{nN}$.

Proof:

If the link cost function $c = c(f)$ is strongly-f-monotone with respect to the constant $a > 0$, then

$$[c(f^1) - c(f^2)]^t[f^1 - f^2] \geq a||c(f^1) - c(f^2)||^2 \quad \forall f^1, f^2 \in K.$$

Making the replacements $f_i = \sum_{p \text{ passing through } i} F_p$, and $C_p(F) = \sum_{i \in \text{path } p} c_i(f)$, and observing that

$$[c(f^1) - c(f^2)]^t[f^1 - f^2] = \sum_{i=1}^{n}[c_i(f^1) - c_i(f^2)][f_i^1 - f_i^2],$$

we obtain

$$[c(f^1)-c(f^2)]^t[f^1-f^2] = \sum_{p=1}^{N}[C_p(F^1)-C_p(F^2)][F_p^1-F_p^2] = [C(F^1)-C(F^2)]^t[F^1-F^2].$$

The defining equality (29) and Lemma 8 imply that

$$\sum_{p=1}^{N}[C_p(F^1) - C_p(F^2)]^2 = \sum_{p=1}^{N}[\sum_{i \in \text{path } p}(c_i(f^1) - c_i(f^2))]^2 \leq$$

$$\leq \sum_{p=1}^{N}(n \sum_{i \in \text{path } p}[c_i(f^1) - c_i(f^2)]^2).$$

Since link i belongs to at most N paths, each term $c_i(f^1) - c_i(f^2)$ appears in the last expression at most N times, so

$$\sum_{p=1}^{N}[C_p(F^1) - C_p(F^2)]^2 \leq nN \sum_{i=1}^{n}[c_i(f^1) - c_i(f^2)]^2.$$

Combining these results shows that

$$[C(F^1) - C(F^2)]^t[F^1 - F^2] = [c(f^1) - c(f^2)]^t[f^1 - f^2] \geq$$

$$\geq a\sum_{i=1}^{n}[c_i(f^1) - c_i(f^2)]^2 \geq \frac{a}{nN}\sum_{p=1}^{N}[C_p(F^1) - C_p(F^2)]^2 = a'||C(F^1) - C(F^2)||^2,$$

if $a' = \frac{a}{nN} > 0$. Therefore, the path cost function $C = C(F)$ is strongly-f-monotone. Q.E.D.

This proposition shows that if the link cost function is strongly-f-monotone, then so is the path cost function. The user optimizing path flow pattern should, therefore, satisfy the following VIP:
find a feasible path flow $F^{opt} \in K$ for which

$$C(F^{opt})^t(F - F^{opt}) \geq 0 \quad \forall F \in K.$$

Consequently, we can apply a path flow projection algorithm to solve the user optimizing traffic equilibrium problem instead of a link flow one. The main step would be the projection

$$F_{k+1} = Pr_K^G(F_k - \rho G^{-1}C(F_k)),$$

in the space of path flows F. Bertsekas and Gafni [6] discuss the advantages of solving the problem in this space of path flows.

As our prior results show, we can consider networks that contain some uncongested paths.

5 Conclusions and open questions

In this paper we introduced an averaging scheme for solving fixed point problems. We consider sequences of averages of the"Riesz" type. The convergence theorem we established involves nonexpansive maps rather than contractive maps. In contrast to the averaging scheme of Baillon [3], at each iteration, our approach uses as the current iterates the average of all prior iterates. Using both Baillon's and our approaches, we established the convergence of the averages of sequences induced by projection and relaxation schemes for variational inequalities, and the generalized steepest descent method for systems of equations. Under a norm condition weaker than an existing one from the literature, we first established the convergence of sequences of averages induced by relaxation algorithms. Assuming the strong-f-monotonicity condition, we showed that averaging schemes induced by the projection algorithm converge to a solution of the variational inequality problem. Finally, assuming the positive semidefinite of the squared Jacobian matrix, we established the convergence of sequences of averages induced by the generalized steepest descent method. We have also shown the connection between strong-f-monotonicity and the norm condition of Pang and Chan [41]. Finally, we applied these results to transportation networks, permitting uncongested links. We showed that whenever the link cost function is strongly-f-monotone then so is the path cost function.

The results in this paper suggest the following questions:

Can we provide convergence results for the sequence of averages of VIP algorithms under weaker conditions?

Can some form of the strong-f-monotonicity condition imposed upon the problem function f guarantee convergence of the sequence of averages induced by other VIP algorithms, such as linearization algorithms and more general iterative schemes?

References

[1] B.H. Ahn and W.W. Hogan (1982), "On the convergence of the PIES algorithm for computing equilibria", *Operations Research*, 30, 281-300.

[2] A. Auslender (1976), *Optimisation: Methodes Numeriques*, Mason, Paris.

[3] J.B. Baillon (1975), "Un theoreme de type ergodique pour les contractions non lineaires dans un espace de Hilbert", *C.R. Acad. Sc., Serie A*, Paris, t.28, 1511-1514.

[4] A.B. Bakushinskij and B.T. Poljak (1974), "On the solution of variational inequalities", *Soviet Mathematics Doklady*, 219, 1705-1710.

[5] D.P. Bertsekas (1982), *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York.

[6] D.P. Bertsekas and E.M. Gafni (1982), "Projection methods for variational inequalities with application to traffic assignment problem", *Mathematical Programming Study*, 17, 139-159.

[7] D.P. Bertsekas and J.N. Tsitsiklis (1989), *Parallel and distributed computation — numerical methods*, Prentice Hall.

[8] R. E. Bruck, Jr. (1977), "On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in hilbert space", *Journal of Mathematical Analysis and Applications*, 61, 159-164.

[9] S. Dafermos (1980), "Traffic equilibria and variational inequalities", *Transportation Science*, 14, 42-54.

[10] S. Dafermos (1982), "Relaxation algorithms for the general asymmetric traffic equilibrium problem", *Transportation Science*, 16, 231-240.

[11] S. Dafermos (1983), "An iterative scheme for variational inequalities", *Mathematical Programming*, 26, 40-47.

[12] J. Eckstein and D. P. Bertsekas (1992), "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators", *Mathematical Programming*, 55, 293-318.

[13] M. Florian and D. Hearn, "Network Equilibria", *Handbook of Operations Research and Management Science* , volume *Network* (M. Ball, T. Magnanti, C. Monma and G. Neumhauser eds.), North-Holland, Amsterdam, to appear.

[14] D. Gabay (1983), "Applications of the method of multipliers to variational inequalities", in M. Fortin, R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, North-Holland, Amsterdam.

[15] A.A. Goldstein (1964), "Convex programming in Hilbert space", *Bull. Amer. Math. Soc.*, vol. 70, no. 5, 709-710.

[16] J.H. Hammond (1984), *Solving Asymmetric Variational Inequality Problems and Systems of Equations with Generalized Nonlinear Programming Algorithms*, Ph.D. dissertation, Department of Mathematics, M.I.T., Cambridge M.A.

[17] J.H. Hammond and T.L. Magnanti (1987), "Generalized descent methods for asymmetric systems of equations", *Mathematics of Operations Research*, 12, 678-699.

[18] P.T. Harker (1993), "Lectures on computation of equilibria with equation-based methods, *CORE Lecture Series.*

[19] P.T. Harker and J-S Pang (1990), "Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications", *Mathematical Programming*, 48, 161-220.

[20] D. Kinderlehrer and G. Stampacchia (1980), *An introduction to variational inequalities and their application*, Academic Press, New York.

[21] K. Knopp (1947), *Theory and application of infinite series*, Hafner Publishing Company N.Y.

[22] A.N. Kolmogorov and S.V. Fomin (1970), *Introduction to Real Analysis*, Dover Publications, New York.

[23] G.M. Korpelevich (1976), "The extragradient method for finding saddle points and other problems", *Economics and Mathematical Methods*, 12, 747-756.

[24] E.S. Levitin and B.T. Polyak (1966), "Constrained minimization methods", *USSR Computational mathematics and mathematical physics*, 6, 1-50.

[25] P.L. Lions and B. Mercier (1979), "Splitting algorithms for the sum of two nonlinear operators", *SIAM Journal on Numerical Analysis*, 16, 964-979.

[26] T.L. Magnanti (1984), "Models and algorithms for predicting urban traffic equilibria", in: M. Florian, ed., *Transportation Planning Models*, North Holland, Amsterdam, 153-185.

[27] T.L. Magnanti and G. Perakis (1993), "A unifying geometric solution framework and complexity analysis for variational inequalities", Working paper OR 276-93, Operations Research Center, M.I.T.

[28] T.L. Magnanti and G. Perakis (1993), "On the convergence of classical variational inequality algorithms", Working paper OR 280-93, Operations Research Center, M.I.T.

[29] T.L. Magnanti and G. Perakis (1993), "Convergence conditions for variational inequality algorithms", Working paper OR 282-93, Operations Research Center, M.I.T.

[30] P. Marcotte (1991), "Application of Khobotov's algorithm to variational inequalities and network equilibrium problems", *INFOR*, 29, 1051-1065.

[31] P. Marcotte and D. Zhu (1993), "Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities", Centre de Recherche sur les Transports, Universite de Montreal, preprint.

[32] O. Mancino and G. Stampacchia (1972), "Convex programming and variational inequalities", *Journal of Optimization Theory and Application*, 9, 3-23.

[33] A. Nagurney (1993), *Network economics: a variational inequality approach*, Kluwer Academic Publishers.

[34] Z. Opial (1967), "Weak convergence of the successive approximation for nonexpansive mappings in Banach Spaces", *Bulletin A.M.S.*, 73, 123-135.

[35] J.M. Ortega and W.C. Rheinboldt (1970), *Iterative solution of nonlinear equations in several variables*, Academic Press , New York.

[36] J.S. Pang (1994), "Complementarity problems", *Handbook of global optimization*, eds. R. Horst and P. Pardalos, Kluwer Academic Publishers.

[37] J.S. Pang (1990), "Newton's method for B-differentiable equations" *Mathematics of Operations Research* 15, 311-341.

[38] J.S. Pang (1990), "Solution differentiability and continuation of Newton's method for variational inequality problems over polyhedral sets", *Journal of Optimization Theory and Application* 66, 121-135.

[39] J.S. Pang (1991), "A B-differentiable, equation-based, globally and locally quadratic convergent algorithm for nonlinear programs, complementarity and variational inequality problems", *Mathematical Programming* 51, 101-131.

[40] J.S. Pang and L. Qi (1993), "Nonsmooth equations: motivation and algorithms", *SIAM Journal on Optimization*, Vol 3.

[41] J.S. Pang and D. Chan (1982), "Iterative methods for variational and complementarity problems", *Mathematical Programming*, 24, 284-313.

[42] G. B. Passty (1979), "Ergodic convergence to a zero of the sum of monotone operators in hilbert space", *Journal of Mathematical Analysis and Applications*, 72, 383-390.

[43] G. Perakis (1992), *Geometric, interior point, and classical methods for solving finite dimensional variational inequality problems*, Ph.D. dissertation, Department of Applied Mathematics, Brown University, Providence R.I.

[44] B.T. Polyak (1963), "Gradient methods for the minimization of functionals", *USSR J. Comput. Math. Math. Phys.*, vol 3, 643-653.

[45] R.T. Rockafellar (1976), "Monotone operators and the proximal algorithm", *SIAM Journal on Control and Optimization*, 14, 877-898.

[46] M. Sibony (1970), "Methodes iteratives pour les equations et inequations aux derivees partielles nonlinearies de type monotone", *Calcolo*, 7, 65-183.

[47] M.J. Smith (1979), "The existence, uniqueness and stability of traffic equilibria", *Transportation Research*, 13B, 295-303.

[48] M.J. Smith (1983), "The existence and calculation of traffic equilibria", *Transportation Research*, 17B, 291-303.

[49] P. Tseng (1990), "Further applications of a matrix splitting algorithm to decomposition in variational inequalities and convex programming", *Mathematical Programming*, 48, 249-264.

[50] J.G. Wardrop (1952), "Some theoretical aspects of road traffic research", *Proc. Inst. Civil Engineers Part II*, 1, 325-378.

[51] E. Zeidler (1990), *Nonlinear functional analysis and its applications II/B: Nonlinear monotone operators*, Springer-Verlag, New York, NY.