Convergence Conditions for Variational Inequality Algorithms

Thomas L. Magnanti *Georgia Perakis*

OR 282-93 October 1993

Convergence Conditions for Variational Inequality Algorithms

Thomas L. Magnanti *　　Georgia Perakis †

October 1993

Abstract

Within the extensive variational inequality literature, researchers have developed many algorithms. Depending upon the problem setting, these algorithms ensure the convergence of (i) the entire sequence of iterates, (ii) a subsequence of the iterates, or (iii) averages of the iterates. To establish these convergence results, the literature repeatedly invokes several basic convergence theorems. In this paper, we review these theorems and a few convergence results they imply, and introduce a new result, called the orthogonality theorem, for establishing the convergence of several algorithms for solving a certain class of variational inequalities. Several of the convergence results impose a condition of strong-f-monotonicity on the problem function. We also provide a general overview of the properties of strong-f-monotonicity, including some new results (for example, the relationship between strong-f-monotonicity and convexity).

Subject Classification:

Programming, Nonlinear, Theory: Convergence Conditions for Variational Inequalities.

Mathematics, Functions: Strongly-f-monotone.

---

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

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

1 Introduction

We consider the variational inequality problem

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

defined over a closed, convex (constraint) set K in R^n. In this formulation $f : K \subseteq R^n \rightarrow R^n$ is a given function and x^{opt} denotes an (optimal) solution of the problem. Variational inequality theory provides a natural framework for unifying the treatment of equilibrium problems encountered in problem areas as diverse as economics, game theory, transportation science, and regional science. Variational inequality problems also encompass a wide range of generic problem areas including mathematical optimization problems, complementarity problems, and fixed point problems.

The literature contains many algorithms for solving variational inequality problems. The convergence results for these algorithms involve the entire sequence of iterates (e.g., [25], [7]), some subsequence of iterates (e.g., [22], [17]), or the sequence of averages of the iterates (e.g., [26], [19]). The review articles by Harker and Pang [13], and by Florian and Hearn [9], the Ph.D. thesis of Hammond [11], and the recent book by Nagurney [23] provide insightful surveys of numerous convergence results and citations to many references in the literature.

Banach's fixed point theorem has been a standard convergence theorem for establishing the convergence of algorithms in many problem settings, including variational inequalities. Two other results, Baillon's Theorem [2] (see [19]), and Opial's Lemma [24] (see [10], [29], [21]) have also proven to be useful tools for establishing convergence results for variational inequalities. In this paper, we briefly summarize the use of these convergence conditions and we introduce a new convergence theorem, the orthogonality theorem. This theorem states that under certain conditions, whenever the map f at an accumulation point of the sequence induced by an algorithm is orthogonal to the line segment between that accumulation point and some variational inequality solution, then every accumulation point of that algorithm is a variational inequality solution. Moreover, if the algorithm map is nonexpansive around a solution for some appropriately defined potential L, then the entire sequence converges to a solution. As part of our discussion, we establish a relationship between the orthogonality theorem and Opial's Lemma.

Some recent convergence results (see for example [10], [29], [17], [19], [21]) impose the condition of strong-f-monotonicity on the problem function f. These results and those in this paper suggest a natural question, *what are the characteristics of strongly-f-monotone functions?* To help answer this question, we provide a general overview of the properties of strong-f-monotonicity, including some new results (for example, the relationship between convexity, or monotonicity in the general asymmetric case, and some form of strong-f-monotonicity).

The remainder of this paper is organized as follows. In Section 2, we provide an overview of several convergence theorems, some algorithms that use them, and the conditions required for these algorithms. We also introduce and prove the orthogonality theorem and relate it to Opial's Lemma. In Section 3, we summarize several properties the strong-f-monotonicity, and introduce some new properties.

2 The Orthogonality Theorem

Banach's celebrated fixed point theorem is a classical result that has been extensively used in the literature to establish the convergence of many algorithms, including those for solving variational inequality problems. The basic condition in this theorem is that the algorithm map is a contraction. Other convergence theorems, which we summarize below, have also proven to be important tools to researchers in establishing various convergence results, and improving on the conditions they need to impose. To state these convergence results, we will impose several conditions on the underlying problem map f and the algorithm map T. If G is a given positive definite matrix, then we define $||x||_G^2 = x^t G x$.

a. A map T is **a contraction map on** K, relative to the $||.||_G$ norm, if,
$||T(x) - T(y)||_G^2 \leq c||x - y||_G^2 \quad \forall x, y \in K$, for some contraction constant $0 < c < 1$.

The use of this condition usually requires the following condition of *strong monotonicty* on the problem function f:

$$\text{for some contant } b > 0, \quad [f(x) - f(y)]^t[x - y] \geq b||x - y||^2 \quad \forall x, y \in K.$$

b. Other convergence results involve nonexpansive estimates. A map T is a **nonexpansive map on** K, relative to the $||.||_G$ norm, if $||T(x) - T(y)||_G^2 \leq ||x - y||_G^2 \ \ \forall x, y \in K$.
The use of this condition on the algorithm map T usually requires the following condition of *strong-f-monotonicity* on the problem function f:

$$\text{for some contant } a > 0, \quad [f(x) - f(y)]^t[x - y] \geq a||f(x) - f(y)||^2 \quad \forall x, y \in K.$$

c. Furthermore, researchers often require the following condition of ordinary *monotonicity* on the problem function f:

$$[f(x) - f(y)]^t[x - y] \geq 0 \quad \forall x, y \in K.$$

Contraction, nonexpansiveness, monotonicity and strong monotonicity are standard conditions in the literature. Gabay [10], implicitly introduced the concept of strong-f-monotonicity and Tseng, [29], using the name co-coercivity, explicitly stated this condition. Magnanti and Perakis ([17], [19], [27] and [18]) have used the term strong-f-monotonicity for this condition, a choice of terminology that highlights the similarity between this concept and the terminology strong monotonicity, which has become so popular in the literature.

Researchers have established convergence properties for variational inequality algorithms using the following basic theorems.

1. Banach's Fixed point Theorem (Banach [14])

Let T be a map, $T : K \rightarrow K$, defined on a closed and convex subset K of R^n. If T is a *contraction* map on K relative to the $||.||_G$ norm, then for every point $y \in K$, the map $T^k(y)$, converges to a fixed point of map T.

2. Baillon's Theorem (Baillon [2])

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

3. Opial's Theorem (Opial [24])

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.

4. In [27] and in subsequent publications, we have established the convergence of new ([17], [20]) and some classical ([19], [20]) algorithms by implicitly using a common proof technique. The following convergence result, which we call the orthogonality theorem, summarizes this proof technique.

The Orthogonality Theorem (see also [18])

Consider the variational inequality problem $VI(f, K)$. Let T be a mapping, $T : K \rightarrow K$, defined over a closed and convex subset K of R^n. Assume that the problem function f and the map T satisfy the following conditions,

(a) The *orthogonality* condition along a subsequence $\{T^{k_j}(y)\} \subseteq \{T^k(y)\}$ for a given point $y \in K$, i.e.,

$$f(T^{k_j}(y))^t(T^{k_j}(y) - x^{opt}) \longrightarrow_{k_j\rightarrow\infty} 0,$$

for some variational inequality solution x^{opt}.

(b) The problem function f is *strongly-f-monotone.*

I. Then every accumulation point of the subsequence $T^{k_j}(y)$ is a variational inequality solution.

II. Furthermore, if for every variational inequality solution x^{opt}, some real-valued potential function $L(x, x^{opt})$ satisfies the conditions, $|L(x, x^{opt})| \geq d||x - x^{opt}||^2$ for some constant $d > 0$ and $L(x^{opt}, x^{opt}) = 0$,

and the map T is *nonexpansive* relative to L around x^{opt}, in the sense that

$$|L(T^{k+1}(y), x^{opt})| \leq |L(T^k(y), x^{opt})|,$$

then the entire sequence $\{T^k(y)\}_{k=0}^{\infty}$ is bounded and converges to a variational inequality solution.

Proof: **I.** We intend to show that, under the assumptions of this theorem, that

$\lim_{k_j\to\infty} f(T^{k_j}(y))^t(x - T^{k_j}(y))$ exists and $\lim_{k_j\to\infty} f(T^{k_j}(y))^t(x - T^{k_j}(y)) \geq 0 \ \ \forall x \in K$. Let x^{opt} be a variational inequality solution. This definition of x^{opt} and the strong-f-monotonicity condition imply that

$$f(T^{k_j}(y))^t(T^{k_j}(y) - x^{opt}) = [f(T^{k_j}(y)) - f(x^{opt})]^t(T^{k_j}(y) - x^{opt}) + f(x^{opt})^t(T^{k_j}(y) - x^{opt}) \geq$$

$$[f(T^{k_j}(y)) - f(x^{opt})]^t(T^{k_j}(y) - x^{opt}) \geq a\|f(x^{opt}) - f(T^{k_j}(y))\|_2^2 \geq 0.$$

The orthogonality condition implies that the left-hand side of these inequalities approaches zero as $k_j \to \infty$. Therefore,

$$\|f(x^{opt}) - f(T^{k_j}(y))\|_2^2 \longrightarrow_{k_j\to\infty} 0 \ \text{ and so } \ f(T^{k_j}(y)) \longrightarrow_{k_j\to\infty} f(x^{opt}). \qquad (2)$$

This result, together with the orthogonality condition, implies that

$$f(T^{k_j}(y))^t T^{k_j}(y) \longrightarrow_{k_j\to\infty} f(x^{opt})^t x^{opt}.$$

But then, for all $x \in K$,

$$\lim_{k_j\to\infty} f(T^{k_j}(y))^t(x - T^{k_j}(y)) = f(x^{opt})^t(x - x^{opt}) \geq 0 \qquad (3)$$

since x^{opt} is a variational inequality solution.

But (3) implies that every accumulation point x^* of the algorithm subsequence $\{T^{k_j}(y)\}$ is indeed a variational inequality solution since $x^* \in K$ and

$$\lim_{k_j\to\infty} f(T^{k_j}(y))^t(x - T^{k_j}(y)) = f(x^*)^t(x - x^*) \geq 0 \quad \forall x \in K.$$

II. Furthermore, if some potential $L(x, x^{opt})$ satisfies the condition $|L(x, x^{opt})| \geq d\|x - x^{opt}\|^2$, $d > 0$, with $L(x^{opt}, x^{opt}) = 0$ and the map T is nonexpansive relative to L around every variational inequality solution x^{opt}, $|L(T^{k+1}(y), x^{opt})| \leq |L(T^k(y), x^{opt})|$, then the monotone sequence $\{L(T^k(y), x^{opt})\}$ is convergent for every solution x^{opt}. Moreover, the entire sequence $\{T^k(y)\}$ is bounded and therefore it has at least one accumulation point. Nevertheless, we have just shown that every accumulation point x^* of the subsequence

$\{T^{k_j}(y)\}$ is a variational inequality solution. Therefore, for that solution x^* the potential converges to $L(x^*, x^*) = 0$ and therefore,

$$0 \leq d||T^k(y) - x^*||^2 \leq |L(T^k(y), x^*)| \longrightarrow_{k\to\infty} 0$$

which implies that the entire sequence $T^k(y)$ converges to that solution x^*. □

Remarks:

1. $L(x, x^{opt}) = ||x - x^{opt}||_2^2$ is an example of a potential that is often used to obtain convergence results. In this case, $L(x^{opt}, x^{opt}) = ||x^{opt} - x^{opt}||_2^2 = 0$ and since $|L(x, x^{opt})| = ||x - x^{opt}||_2^2$, $d = 1$. For this potential function, requiring the map T to be nonexpansive relative to L around every solution is the same as requiring the map T to be nonexpansive, relative to the $||.||$ norm, around every solution. $L(x, x^{opt}) = K(x^{opt}) - K(x) - K'(x)^t(x^{opt} - x)$ is another example of a potential, which appears in Cohen's auxiliary problem framework (see remark (b) below and [4] and [21] for more details on this framework).

2. In the symmetric Frank-Wolfe algorithm, if we impose the strong-f-monotonicity condition on the problem function f, then part (I) of the orthogonality theorem holds, and so every accumulation point of the algorithm sequence is a solution (see [19] for more details). If we replace strong-f-monotonicity with strong monotonicity, however, then part (II) of the theorem also holds, with a potential $L(x, x^{opt}) = F(x) - F(x^{opt})$, when F is the objective function of the minimization problem corresponding to the variational inequality problem.

3. The orthogonality theorem seems to suggest a new potential, the potential involved in the orthogonality condition, that might be used to obtain convergence results for variational inequality problems. In [20], we have used this potential to establish the convergence of a descent framework for solving variational inequalities, which includes as special cases the steepest descent method, the Frank-Wolfe method (symmetric and asymmetric), linearization schemes [25], the generalized contracting ellipsoid method [11], and others.

Table I summarizes the use of the four basic convergence theorems we have introduced for establishing the convergence of various variational inequality algorithms.

Figure I illustrates the orthogonality theorem when the algorithm sequence has two accumulation points x^* and x^{**}, which are both VIP solutions.

To provide further insight concerning the orthogonality theorem, we now establish a relationship between the orthogonality condition and the asymptotic regularity condition in Opial's Lemma. As we show, for variational inequalities, Opial's assumptions imply those imposed in the orthogonality theorem. The opposite is also true for the general iterative scheme [7] of Dafermos if we impose some additional assumptions. We will also show, through a counterexample, that in general the opposite is not true. Therefore, the orthogonality theorem is more general than Opial's Lemma.

Proposition 2.1:

Let $VI(f, K)$ be a variational inequality problem with a monotone problem function f. Consider a mapping $T : K \rightarrow K$ satisfying the property that every fixed point of this mapping is a variational inequality solution. Then for every $y \in K$ the asymptotic regularity condition of T, i.e., $||T^{k+1}(y) - T^k(y)||_2 \longrightarrow_{k\rightarrow\infty} 0$,

implies the orthogonality condition along any convergent subsequence $\{T^{k_j}(y)\} \subseteq \{T^k(y)\}$, i.e.,

$$f(T^{k_j}(y))^t(T^{k_j}(y) - x^{opt}) \longrightarrow_{k_j\rightarrow\infty} 0.$$

Proof: When $||T^{k+1}(y) - T^k(y)||_2 \longrightarrow_{k\rightarrow\infty} 0$ for some $y \in K$, then every accumulation point x^* of the sequence $\{T^k(y)\}$ is a fixed point of the mapping T and, therefore, by assumption x^* is a variational inequality solution. But since x^* solves the variational inequality problem, the monotonicity of f implies that for any optimal solution x^{opt}, $0 \geq f(x^*)^t(x^* - x^{opt}) \geq 0$, and, therefore,

for any convergent subsequence $\{T^{k_j}(y)\}$, $f(T^{k_j}(y))^t(T^{k_j}(y) - x^{opt}) \longrightarrow_{k_j\rightarrow\infty} 0$. □

When is the converse true? We will answer this question for the case of a general iterative scheme. In fact, we will show that in the case of the general iterative scheme, under some additional assumptions, the asymptotic regularity condition is equivalent to the orthogonality condition. The general iterative scheme (see [7]) determines the point x^{k+1} from the previous iterate x^k by solving the variational inequality,

$$\text{find } x^{k+1} \in K \text{ satisfying, } g(x^{k+1}, x^k)^t(z - x^{k+1}) \geq 0 \quad \forall z \in K, \tag{4}$$

assuming that $g(x, y)$ is defined so that $g(x, x) = \rho f(x)$ for some constant $\rho > 0$.

Proposition 2.2:

Let $VI(f, K)$ be a variational inequality problem. Consider the general iterative scheme (4). Let $T : K \longrightarrow K$ be a function that maps a point $y = x^k$ into a point $T(y) = x^{k+1}$ that solves (4). If the problem function f is monotone, $||x - x^{opt}||$ is uniformly bounded over $x \in K$, and some constant $C > 0$ satisfies the condition $||\nabla_y g(x, y)|| \leq C$ for all $x, y \in K$, then the asymptotic regularity condition on the map T implies the orthogonality condition across the *entire* sequence. Conversely, if

1. the problem function f is strongly-f-monotone with constant a,

2. the scheme's function $g(x, y)$ is strongly monotone relative to its x component, i.e.,

for some constant $b > 0$, $[g(x_1, y) - g(x_2, y)]^t[x_1 - x_2] \geq b||x_1 - x_2||^2$, $\forall x_1, x_2 \in K$, and

3. the constant $0 < \rho < 4ab$ (often $\rho = 1$, then this condition requires $1 < 4ab$),

then the orthogonality condition along some subsequence, implies the asymptotic regularity along that subsequence.

3'. Replacing 3 with the assumption that the orthogonality condition holds along the *entire* sequence $\{T^k(y)\}$ (with no restrictions on ρ) also implies the asymptotic regularity condition.

Proof: " $\Rightarrow$" Set $T^k(y) = x^k$ and $T^{k+1}(y) = x^{k+1}$. If T is asymptotically regular, then $||x^k - x^{k+1}|| \longrightarrow_{k \to \infty} 0$. The fact that $g(x, x) = \rho f(x)$, f is monotone, and $x^{opt} \in K$ is a variational inequality solution implies that

$$0 \leq g(x^{k+1}, x^k)^t(x^{opt} - x^{k+1}) \leq [g(x^{k+1}, x^k) - g(x^{k+1}, x^{k+1})]^t[x^{opt} - x^{k+1}] \leq$$

(an application of the mean value and Cauchy's inequality imply that)

$$||x^k - x^{k+1}||.||\nabla_y g(x^{k+1}, y)||.||x^{opt} - x^{k+1}|| \leq C||x^k - x^{k+1}||.||x^{opt} - x^{k+1}|| \longrightarrow_{k \to \infty} 0.$$

$$\text{Therefore, } g(x^{k+1}, x^k)^t(x^{opt} - x^{k+1}) \longrightarrow_{k \to \infty} 0 \text{ and}$$

$$[g(x^{k+1}, x^k) - g(x^{k+1}, x^{k+1})]^t[x^{opt} - x^{k+1}] \longrightarrow_{k \to \infty} 0.$$

So $\rho f(x^{k+1})^t(x^{opt}-x^{k+1}) = [g(x^{k+1},x^{k+1})]^t[x^{opt}-x^{k+1}] =$

$$[g(x^{k+1},x^k)]^t[x^{opt}-x^{k+1}] - [g(x^{k+1},x^k) - g(x^{k+1},x^{k+1})]^t[x^{opt}-x^{k+1}] \longrightarrow_{k\to\infty} 0.$$

Therefore, the orthogonality condition holds, i.e., $f(x^k)^t(x^{opt}-x^k) \longrightarrow_{k\to\infty} 0$.

" $\Leftarrow$" Conversely, (a) assume that the orthogonality condition holds along some subsequence, i.e., $f(x^{k_j})^t(x^{opt}-x^{k_j}) \longrightarrow_{k_j\to\infty} 0$. Then the general iterative scheme and the fact that $x^{k_j}\in K$ imply that

$$0 \le g(x^{k_j+1},x^{k_j})^t(x^{k_j}-x^{k_j+1}) = [g(x^{k_j+1},x^{k_j}) - g(x^{k_j},x^{k_j})]^t(x^{k_j}-x^{k_j+1}) +$$

$$g(x^{k_j},x^{k_j})^t(x^{k_j}-x^{k_j+1}) \le$$

(the strong monotonicity of $g(x,y)$ relative to x, and the fact that $g(x,x)=\rho f(x)$ implies that)

$$-b||x^{k_j+1}-x^{k_j}||^2 + \rho f(x^{k_j})^t(x^{k_j}-x^{k_j+1}) = -b||x^{k_j+1}-x^{k_j}||^2 +$$

$$\rho f(x^{k_j})^t(x^{k_j}-x^{opt}) + \rho f(x^{k_j})^t(x^{opt}-x^{k_j+1}) \le$$

(the definition of a VIP solution x^{opt} implies that)

$$-b||x^{k_j+1}-x^{k_j}||^2 + \rho f(x^{k_j})^t(x^{k_j}-x^{opt}) +$$

$$\rho[f(x^{k_j}) - f(x^{opt})]^t(x^{opt}-x^{k_j+1}) = -b||x^{k_j+1}-x^{k_j}||^2 + \rho f(x^{k_j})^t(x^{k_j}-x^{opt}) +$$

$$\rho[f(x^{k_j}) - f(x^{opt})]^t(x^{opt}-x^{k_j}) + \rho[f(x^{k_j}) - f(x^{opt})]^t(x^{k_j}-x^{k_j+1}) \le$$

(strong-f-monotonicity implies that)

$$-b||x^{k_j+1}-x^{k_j}||^2 + \rho f(x^{k_j})^t(x^{k_j}-x^{opt}) - a\rho||f(x^{k_j}) - f(x^{opt})||^2 + \rho[f(x^{k_j}) - f(x^{opt})]^t(x^{k_j}-x^{k_j+1}) \le$$

(by expanding $||f(x^{k_j}) - f(x^{opt}) + \frac{1}{2\sqrt{a}}(x^{k_j}-x^{k_{j+1}})||^2$ and rearranging terms we obtain)

$$[-b+\frac{\rho}{4a}]||x^{k_j+1}-x^{k_j}||^2 + \rho f(x^{k_j})^t(x^{k_j}-x^{opt}).$$

If $0<\rho<4ab$, then $0<[b-\frac{\rho}{4a}]||x^{k_j+1}-x^{k_j}||^2 \le \rho f(x^{k_j})^t(x^{k_j}-x^{opt})$.

Therefore, the orthogonality condition along the subsequence x^{k_j} implies the asymptotic regularity along that subsequence.

(b) Assume that there are no restrictions on ρ and the orthogonality condition holds along the entire sequence, i.e., $f(x^k)^t(x^{opt} - x^k) \longrightarrow_{k\to\infty} 0$. The same argument that led us to (2) in part (I) of the orthogonality theorem implies that $f(x^k) \longrightarrow_{k\to\infty} f(x^{opt})$. Furthermore, as in part (a), we find that

$$0 \leq -b\|x^{k+1}-x^k\|^2+\rho f(x^k)^t(x^k-x^{opt})+\rho[f(x^k)-f(x^{k+1})]^t(x^{opt}-x^{k+1})+\rho f(x^{k+1})^t(x^{opt}-x^{k+1}).$$

Therefore

$$b\|x^{k+1}-x^k\|^2 \leq \rho f(x^k)^t(x^k-x^{opt})-\rho f(x^{k+1})^t(x^{k+1}-x^{opt})+\rho[f(x^k)-f(x^{k+1})]^t(x^{opt}-x^{k+1}).$$

Cauchy's inequality implies that

$$b\|x^{k+1} - x^k\|^2 \leq \rho f(x^k)^t(x^k - x^{opt}) - \rho f(x^{k+1})^t(x^{k+1} - x^{opt})+$$

$$\rho\|f(x^k) - f(x^{k+1})\|.\|x^{opt} - x^{k+1}\| \longrightarrow_{k\to\infty} 0.$$

Therefore, asymptotic regularity holds: $\|x^{k+1} - x^k\| \longrightarrow_{k\to\infty} 0$. Notice that part (b) holds regardless of the choice of ρ. □

Remarks:

1. Several classical methods for solving the variational inequality problem, which are special cases of the general iterative scheme, satisfy the strong monotonicity condition on the scheme's function g if we impose some conditions on the underlying data.

   (a) Linear approximation methods ([25]) with $g(x, y) = f(y) + A(y)^t(x - y)$. In this case the matrix $A(y)$ should be uniformly positive definite for all y and the problem function f strongly-f-monotone.

   The following examples are special cases of this class of algorithms.

   The Linearized Jacobi method, with $A(y) = diag(\nabla f(y))$, which should have positive elements.

   The Projection method with $A(y) = G$, a positive definite matrix.

   Newton's method with $A(y) = \nabla f(y)$, a uniformly positive definite matrix.

The Quasi-Newton method with $A(y) = approx(\nabla f(y))$, which we require to be a uniformly positive definite matrix.

The Linearized Gauss-Seidel method with $A(y) = L(y) + D(y)$, or $A(y) = U(y) + D(y)$ ($L(y)$ and $U(y)$ are the lower and upper diagonal parts of the matrix $\nabla f(y)$). $A(y)$ should be a uniformly positive definite matrix.

(b) For **Cohen's auxiliary problem framework** ([4], [21]) with $g(x,y) = \rho f(y) + G(x) - G(y)$. In this case, the problem function f should be strongly-f-monotone and the function G should be strongly monotone (in fact, Cohen assumes that $G(y) = K'(y)$, that is a gradient matrix for a strongly convex funcion K).

2. As shown by the following example, the orthogonality theorem is more general than Opial's lemma. Consider the symmetric Frank-Wolfe algorithm (for more details on the Frank-Wolfe algorithm see, for example, [22] and [11]) with problem iterates x^k: At each step $k = 1, 2, ...,$ solve the linear program $y^k = argmin_{x \in K} f(x^{k-1})^t x$ and then solve the following 1-dimensional variational inequality problem, i.e., find $x^k \in [y^k; x^{k-1}]$ satisfying $f(x^k)^t(x - x^k) \geq 0, \quad \forall x \in [y^k; x^{k-1}]$.

   As we have shown in [19], the orthogonality theorem implies that this algorithm converges to a VIP solution along a subsequence. As shown by the following example, for this algorithm the orthogonality condition need not always satisfy the asymptotic regularity condition.

   Example: Let $K = \{x = (x_1, x_2) \in R^2 : \quad 0 \leq x_1 \leq 1, \ 0 \leq x_2 \leq 1\}$ be the feasible set

   and let $f(x) = \left\{ \begin{array}{lll} (x_1 - \frac{1}{4}, 1) & \text{if} & 0 \leq x_1 \leq \frac{1}{4} \\ (0,1) & \text{if} & \frac{1}{4} \leq x_1 \leq \frac{3}{4} \\ (x_1 - \frac{3}{4}, 1) & \text{if} & \frac{3}{4} \leq x_1 \leq 1 \end{array} \right\}$ be the problem function f.

   It is easy to see that f is a Lipschitz continuous and a strongly-f-monotone function (but not strict or strongly monotone) with a symmetric Jacobian matrix. The VIP solutions of this problem are all the points $x^{opt} = (x_1^{opt}, 0)$ with $\frac{1}{4} \leq x_1^{opt} \leq \frac{3}{4}$. If we start the algorithm at the point $x^0 = (0, \frac{1}{8})$, then

   step k=1 solves at $y^1 = (1,0)$ and $x^1 = (\frac{7}{8}, \frac{1}{64})$.

Step k=2 solves at $y^2 = (0,0)$ and $x^2 = \frac{13}{49}(\frac{7}{8}, \frac{1}{64})$. Starting at the point $x^0 = (0, \frac{1}{8})$, (or any point $(0, z)$ with $z \leq \frac{1}{8}$) the algorithm induces two subsequences, $\{x^{2l-1}\}_{l=0}^{\infty}$ and $\{x^{2l}\}_{l=0}^{\infty}$ with two accumulation points $x^* = (\frac{3}{4}, 0)$ and $x^{**} = (\frac{1}{4}, 0)$ that are both VIP solutions. The asymptotic regularity condition and hence Opial's Lemma does not hold since $||x^k - x^{k+1}|| \longrightarrow_{k\to\infty} \frac{1}{2}$. It is easy to check, the orthogonality condition holds for both subsequences so in this example the orthogonality theorem applies, but not Opial's lemma. Figure I illustrates this example.

3. Table I illustrates the use of various convergence conditions for solving variational inequalities. As indicated in this table, the use of the orthogonality theorem establishes the convergence of several algorithms, whose convergence has not been established using Opial's Lemma. These algorithms include the general geometric framework [17], the Frank-Wolfe algorithm [19], and a descent framework [20].

3 On the strong-f-monotonicity condition

As shown in the previous section, and particularly as summarized in Table I, the strong-f-monotonicity condition plays an important role as an underlying condition for establishing the convergence of several algorithms. For example, as we have seen, the orthogonality theorem requires that a problem map f that is strongly-f-monotone.

In this section we provide a general overview of several properties for characterizing strong-f-monotonicity. We begin by reviewing several properties that Magnanti and Perakis ([17], [19], [18], [27]) and Marcotte and Zhu ([21]) have independently established.

3.1 Some known results

Proposition 3.1: ([17])

The problem function f is strongly-f-monotone if and only if its generalized inverse f^{-1} is strongly monotone in $f(K)$.

This result is an immediate consequence of the definition of strong-f-monotonicity and the generalized inverse.

Table II summarizes five types of monotonicity and their differential conditions (see [17], [19] for more details). Whenever the problem function f satisfies any one of these differential conditions, then f also satisfies the corresponding monotonicity conditions.

Proposition 3.2: ([17])

For affine functions f (i.e., $f(x) = Mx - c$, for some matrix M and vector c), the differential form of strong-f-monotonicity holds if we can

$$\text{find a constant } a > 0 \text{ so that } M^t - aM^tM \text{ is a positive semidefinite matrix.}$$

(Note that constant functions $f(x) = c$ for all x satisfy this condition.)

Remark: One of the principal attractions of strongly-f-monotone functions is the fact that the class of variational inequalities with strongly-f-monotone functions contains all linear programs, (when the feasible set K is a polyhedron, $f(x) = c$). Recall that linear programs do not always have optimal solutions (since the defining polyhedron might be unbounded), and so variational inequalities with strongly-f-monotone functions need not have a solution.

Proposition 3.3: ([17], [19])

The following statements are true,

(i) Any strongly-f-monotone function is monotone.

(ii) Suppose f is one-to-one (i.e., invertible) and so by Proposition 3.1, strong-f-monotonicity of f is equivalent to strong monotonicity of the regular inverse f^{-1}.

Then if $f_i(x) \neq f_i(y)$ for some $i = 1, ..., n$, whenever $x \neq y$ (i.e., at least one component f_i of f is invertible), then strong-f-monotonicity implies strict monotonicity.

(iii) The strong-f-monotonicity of f with constant a implies the Lipschitz continuity of f with constant $\frac{1}{a}$ (see also [30]).

(iv) If f is Lipschitz continuous, then strong monotonicity implies strong-f-monotonicity.

Therefore, the class of strongly-f-monotone functions is a class of Lipschitz continuous functions that lies between the classes of monotone functions and strongly monotone functions.

Figure II also summarizes some of the basic properties of strong-f-monotonicity (see also [17], [19]).

Proposition 3.4: ([21])

Consider the variational inequality problem $VI(f,K)$. If $af - I$ is a Lipschitz continuous function over K for a constant ≤ 1, then f is a strongly-f-monotone function on K for the constant $\frac{a}{2}$. Conversely, if f is strongly-f-monotone on K for a constant $\geq \frac{a}{2}$ on K then $af - I$ is also Lipschitz continuous for the constant 1.

The following proposition states a related result.

Proposition 3.5: ([19])

The matrix $M^t[I - \frac{a}{2}M]$ is positive semidefinite for $a > 0$ if and only if the operator norm $||I - aM|| \leq 1$.

M is a general matrix. Note that if $f(x) = Mx - b$, then the results in Proposition 3.4 and 3.5 are the same because the positive semidefinite condition in Proposition 3.5 is the differential form for strong-f-monotonicity. In general, if $M = \nabla f$, then $||I - a\nabla f(z)|| \leq 1$ is a Lipshitz continuity condition of the differential of $Ix - af(x)$, but evaluated at a given point $x = z$ of the map $\nabla f(x)$ (that is, for the fixed matrix $I - a\nabla f(z)$).

In establishing convergence of the general iterative scheme (see (4)) and several of its variations and specializations, researchers (see [7] and [25]) have invoked a norm condition

$$||g_x^{-1/2}(x,x)g_y(x,x)g_x^{-1/2}(x,x)|| < 1 \quad \forall x \in K. \tag{5}$$

As shown by the following proposition, these results provide another setting illustrating the importance of strong-f-monotonicity. In fact, as shown by the following result, the norm condition (5) is equivalent to a weak version of the differential form of strong-f-monotonicity of Table II (with $x = y$).

Proposition 3.6: ([19])

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 $\rho > 0$ is some constant satisfying the condition $\rho \leq 2g_{min}a$ 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 (5) with $\leq$ form instead of $<$).

2. Conversely, if the norm condition (5) holds in a less than or equal to form, then for some constant $0 < a \leq \frac{\rho}{2g_{max}}$, 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 $x \in K$.

Proposition 3.6 also holds for the differential form of strict strong-f-monotonicity. Then the norm condition holds as a strict inequality.

In the prior discussion, $\rho > 0$. Often $\rho = 1$. Then the strong-f-monotonicity constant a should satisfy $a \geq \frac{1}{2g_{min}}$ for part (1) and $0 < a \leq \frac{1}{2g_{max}}$ for part (2).

In some situations, for example the well-known traffic equilibrium problem, it is more efficient to work with a transformed set of variables.

Proposition 3.7: ([19], [18], [21])

In the setting of the traffic equilibrium problem [5], let n be the total number of links in the network and N the total number of paths. Then if the link cost function is strongly-f-monotone for a constant a, then so is the path cost function for the constant $a' = \frac{a}{nN}$.

In general, the sum of strongly-f-monotone functions is also strongly-f-monotone. Moreover, affine transformations preserve strong-f-monotonicity.

Proposition 3.8: ([21])

If the problem function f is strongly-f-monotone for a constant a and A is an $n \times m$ matrix, then the function $A^t f A + c$ is also strongly-f-monotone with the constant $\frac{a}{\|A\|}$.

As a last result in this subsection, we note that if a function satisfies the *weak differential form* of strong-f-monotonicity, i.e., for all $w \in R^n$ and $x_1 \in K$, some constant $a > 0$ satisfies the condition,

$$w^t \nabla f(x_1) w \geq a w^t \nabla f(x_1)^t \nabla f(x_1) w,$$

then $w^t \nabla f(x_1) w = 0$ for any x_1 and w implies that $w^t \nabla f(x_1)^t \nabla f(x_1) w = 0$ and $\nabla f(x_1) w = 0$. Luo and Tseng [16] have studied a class of matrices $B = \nabla f(x_1)$ that satisfy this property.

Definition 1 *: (Luo and Tseng [16]) A matrix B is positive semidefinite plus (p.s.d. plus) if it is positive semidefinite and*

$$\text{if} \quad x^t B x = 0 \quad \text{implies that} \quad Bx = 0.$$

Luo and Tseng [16] have shown that the class of p.s.d. plus matrices B is equivalent to the class of matrices that can be decomposed into the product $B = P^t P_1 P$ for some P_1 positive definite matrix and some (possibly nonsquare) matrix P. Note that every symmetric, positive semidefinite matrix B is p.s.d. plus, since in this case $B = H^t H$ for some matrix H and so $x^t B x = 0$ implies $x^t H^t H x = 0$, which implies that $Hx = 0$ and therefore $Bx = 0$.

3.2 Symmetry, Positive Semidefinite Plus, Uniform Positive Semidefinite Plus, and Positive Semidefiniteness of the Squared Jacobian Matrix

Having reviewed some known properties of strongly-f-monotone functions, we now consider a few new results, some that extend prior results from the literature.

Consider the symmetric case in which $f = \nabla F$ for some twice differentiable function F, and so $\nabla F(x)$ is symmetric for all $x \in K$. In this case, since strong-f-monotonicity implies monotonicity, $\nabla f(x)$ is positive semidefinite for all $x \in K$ (see Table II), and so strong-f-monotonicity implies that F is a convex function.

Is the converse true? That is, does convexity imply strong-f-monotonicity and, if not, how far from convexity need we stray to ensure that the function f is strongly-f-monotone? The following results give a partial answer to this questions. We begin by giving two examples. The first example shows that convexity of F does not imply that f is strongly-f-monotone and the second example shows that even when F is convex over a compact set, it might still not be strongly-f-monotone.

Example 2: Consider the variational inequality with the feasible noncompact set $K = \{x = (x_1, x_2) \in R^2 : \ x_1 \geq 0, \ x_2 \geq 0\}$ and the problem function $f(x) = (x_1^2, 1)$. $f(x) = \nabla F(x)$ with $F(x) = \frac{x_1^3}{3} + x_2$, which is a convex function over K. In this case, f is not strongly-f-monotone on K since for $y = (y_1, y_2)$ and $x = (x_1, x_2)$, with $x_1 = y_1 + 2$, *no* constant $a > 0$ satisfies the condition

$$[f(x) - f(y)]^t [x - y] = 2[2y_1 + 4] \geq a\|f(x) - f(y)\|^2 = [2y_1 + 4]^2, \quad \text{for all} \quad y \in K.$$

Example 3: Consider the variational inequality with the feasible set

$K = \{x = (x_1, x_2) \in R^2 : \ x_1 \leq 1, \ x_2 \leq 1, \ x_1 + x_2 \geq 1, \ x_1 \leq x_2\}$ and problem function $f(x) = (\frac{x_1^2}{2} - \frac{(1-x_2)^2}{2}, -(1-x_1)(1-x_2))$. $\nabla f(x) = \begin{bmatrix} x_1 & 1-x_2 \\ 1-x_2 & 1-x_1 \end{bmatrix}$ is a symmetric, positive semidefinite matrix over K. $F(x) = \frac{x_1^3}{6} + \frac{(1-x_1)(1-x_2)^2}{2}$ is convex in K, since its Hessian matrix is $\nabla^2 F(x) = \nabla f(x) = \begin{bmatrix} x_1 & 1-x_2 \\ 1-x_2 & 1-x_1 \end{bmatrix}$.

For the points $x = (1,1)$ and $y = (\frac{1}{2}, \frac{1}{2})$, $\nabla f(x) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $\nabla f(y) = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

Then for all $w = (w_1, w_2) \in R^2$, $w^t \nabla f(x)^t w = w_1^2$ and $w^t \nabla f(x)^t \nabla f(y) w = \frac{1}{2}(w_1^2 + w_1 w_2)$ which implies that *no* constant $a > 0$ satisfies the condition $w^t \nabla f(x)^t w \geq a w^t \nabla f(x)^t \nabla f(y) w$ for all $w = (w_1, w_2) \in R^2$,

since for $w_1 \longrightarrow 0$ and $w_2 = 1$, $\frac{w^t \nabla f(x)^t \nabla f(y) w}{w^t \nabla f(x)^t w} = \frac{1}{2} + \frac{w_2}{2w_1} \longrightarrow \infty$.

Before continuing, we might note a relationship between the convexity of F and the weak differential form of strong-f-monotonicity.

Proposition 3.9: ([18])

Suppose that $F : K \subseteq R^n \rightarrow R$ is a continuous function and the maximum eigenvalue of the Hessian matrix $\nabla^2 F(x) = \nabla f(x)$ is bounded for all $x \in K$, i.e., if $d_i(x)$ is the ith eigenvalue of $\nabla^2 F(x)$, then $sup_{x \in K}[max_{\{i=1,\ldots,n\}} d_i(x)] \leq d$ for some positive constant d. Then F is convex if and only if for all $x \in K$ and for all $w \in R^n$, $w^t \nabla f(x) w \geq a w^t \nabla f(x)^t \nabla f(x) w$ for some constant $a > 0$.

Proof: "$\Rightarrow$" Assume that the Hessian matrix $\nabla^2 F(x) = \nabla f(x)$ is positive semidefinite for all $x \in K$. Recall from linear algebra that any symmetric, positive semidefinite matrix $M \neq 0$ has an orthogonal representation, i.e., $M = P^t D P$, for some orthogonal matrix P (whose columns are the orthonormal eigenvectors of M). In this representation, D is the diagonal matrix whose elements are the eigenvalues of M (which are nonnegative in the positive semidefinite case). When applied to a symmetric Jacobian matrix, this result implies the existence of an orthogonal matrix $P(x)$ (the operator norm $\|P(x)\| = 1$) satisfying $\nabla f(x) = P(x)^t D(x) P(x)$. Then $w^t \nabla f(x) w = w^t P(x)^t D(x) P(x) w$ and

$w^t \nabla f(x)^t \nabla f(x) w = w^t P(x)^t D(x)^t P(x) P(x)^t D(x) P(x) w = w^t P(x)^t D(x)^2 P(x) w$.

$D(x)$ is a diagonal matrix with diagonal elements $d_i(x)$. Requiring $w^t \nabla f(x) w \geq a w^t \nabla f(x)^t \nabla f(x) w$ for some constant $a > 0$ is equivalent to requiring

$$\sum_i d_i(x)(P(x)w)_i^2 \geq a \sum_i d_i(x)^2 (P(x)w)_i^2.$$

If for all i and $x \in K$, $d_i(x) = 0$ then this inequality is true for any $a > 0$.

If for at least one i and $x \in K$, $d_i(x) \neq 0$, then since $sup_{x \in K}[max_i d_i(x)] \leq d$ and the matrix $\nabla f(x)$ is positive semidefinite, setting $a = \frac{1}{d}$ gives $d_i(x) \geq a d_i(x)^2$ for all i and $x \in K$, which implies the inequality.

" $\Leftarrow$" The converse is easy to see. When the differential form of strong-f-monotonicity holds for all $x_1 = x_2 = x$, then the Jacobian matrix is positive semidefinite and therefore the function F is convex. □

Remark:

Propositions 3.5 and 3.9 show that on a compact set, F is convex (and so the maximum eigenvalue of $\nabla f(x)$ is bounded over K) if and only if $\|I - a\nabla f(x)\| \leq 1$ for all $x \in K$. In terms of the general iterative scheme (4) (see [7], [25]), if $g_x(x,x)$ is positive definite for all $x \in K$, Proposition 3.6 and this result imply that

on a compact set, F is convex if and only if $\|g_x^{-1/2}(x,x) g_y(x,x) g_x^{-1/2}(x,x)\| \leq 1 \ \ \forall x \in K$,

i.e., the less than or equal form of the norm condition holds. This observation shows that the less than or equal form of the norm condition (on a compact set) that researchers have invoked previously in the literature together with symmetry of $\nabla f(x)$ implies convexity. Therefore, we can view the norm condition when applied to asymmetric problems as a form of generalization of convexity.

Although Proposition 3.9 shows a connection between convexity and strong-f-monotonicity, it does not show that convexity implies strong-f-monotonicity since it requires $x_1 = x_2 = x$ in the differential condition.

As our previous examples show, we need to impose additional structure on f or on K to ensure that the convexity of F implies strong-f-monotonicity. Furthermore, we might want to ask, *what happens in the general asymmetric case?"*

First, we make the following observation.

Lemma 3.1: Every p.s.d. plus matrix $M(x)$ can be rewritten as

$M(x) = P(x)^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix} P(x)$, for some $n_1(x) \times n_1(x)$, positive definite matrix $P_0(x)$ and some square matrix $P(x)$. Conversely, any matrix $M(x)$ that is of this form is also p.s.d. plus.

Proof: " $\Rightarrow$" Luo and Tseng [16] have shown that it suffices to show that whenever $w^t M(x) w = 0$, then $M(x)w = 0$.

Suppose $M(x)$ can be written as $M(x) = P(x)^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix} P(x)$. Then

$w^t M(x) w = w^t P(x)^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix} P(x) w = [v^t, y^t, z^t] \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v \\ y \\ z \end{bmatrix} = y^t P_0(x) y =$

0, with $P(x)w = \begin{bmatrix} v \\ y \\ z \end{bmatrix}$ and y a vector that has the same dimension as $P_0(x)$. But since $P_0(x)$ is a positive definite matrix, $y = 0$ and therefore

$P(x)^t \begin{bmatrix} 0 \\ P_0(x)y \\ 0 \end{bmatrix} = P(x)^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix} P(x) w = M(x) w = 0$. Therefore, $M(x)$ is also a p.s.d. plus matrix.

" $\Leftarrow$" Conversely, if an $n \times n$ matrix $M(x)$ is p.s.d. plus, then for some $n_1(x) \times n_1(x)$, positive definite matrix $P_0(x)$ we can rewrite $M(x) = P''(x)^t P_0(x) P''(x)$, with $P''(x)$ an $n_1 \times n$ matrix. Then setting $P(x)^t = [P'(x)^t, P''(x)^t, P'''(x)^t]$ for any matrices $P'(x)$ and $P'''(x)$, with appropriate dimensions, we can conclude that $M(x)$ can be rewritten as $P(x)^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix} P(x)$. □

We next address the following question:

"what is the analog of Proposition 3.9 for the general asymmetric case?"

"how asymmetric can the Jacobian matrix be?"

Proposition 3.10:

The positive semidefiniteness of the Jacobian matrix of the problem function f and of the squared Jacobian matrix implies the weak differential form of the strong-f-monotonicity condition, which in turn implies the positive semidefiniteness of the Jacobian matrix of the problem function f, i.e., ordinary monotonicity.

Proof: First, we make the following observations about a general asymmetric matrix M. $w^tMw = w^t\frac{M+M^t}{2}w$ and, when M^2 is positive semidefinite,

$$||\frac{M+M^t}{2}w||^2 = w^t(\frac{M+M^t}{2})^t(\frac{M+M^t}{2})w =$$
$$w^t\frac{M^2+(M^2)^t+M^tM+MM^t}{4}w \geq \frac{||Mw||^2}{4}.$$

Therefore, the positive semidefiniteness of the Jacobian matrix, the previous observations, and Remark (i) following Proposition 3.9, applied to the symmetric matrix $\frac{\nabla f(x)^t+\nabla f(x)}{2}$, imply that for some constant $a > 0$, and for all x, and w,

$$w^t\nabla f(x)w = w^t\frac{\nabla f(x)+\nabla f(x)^t}{2}w \geq a||\frac{\nabla f(x)+\nabla f(x)^t}{2}w||^2 \geq a||\frac{\nabla f(x)}{2}w||^2 \geq \frac{a}{4}w^t\nabla f(x)^t\nabla f(x)w,$$

which is the weak differential form of the strong-f-monotonicity condition.
Furthermore, the weak differential form of the strong-f-monotonicity condition, i.e.,

there is $a > 0$ satisfying the condition, $w^t\nabla f(x)w \geq aw^t\nabla f(x)^t\nabla f(x)w$ for all x w,

implies the positive semidefiniteness of the Jacobian matrix. □

Hammond and Magnanti [12] originally introduced the condition of positive definiteness of the squared Jacobian matrix while establishing the convergence of the steepest descent method for variational inequalities. This condition implies that the Jacobian matrix cannot be "very" asymmetric. In fact, the squared Jacobian matrix is positive definite when the angle between $\nabla f(x)w$ and $\nabla f(x)^t w$ is less than 90 degrees for all $w \in R^n$.

Proposition 3.11:

The converse of the statements in Proposition 3.10 are not valid.

Proof: To establish this result, we will provide counterexamples.

Example 4: The weak differential form of strong-f-monotonicity does not imply that the square of the Jacobian matrix is positive semidefinite.

Consider the function $f(x) = Mx$ with the Jacobian matrix $M = \begin{bmatrix} c & b \\ -b & c \end{bmatrix}$ and let $0 < c < b$. Then $M^2 = \begin{bmatrix} c^2 - b^2 & 2cb \\ -2cb & c^2 - b^2 \end{bmatrix}$ is a negative definite matrix, which means that M^2 is not positive semidefinite. Nevertheless the function $f(x) = Mx$ is strongly-f-monotone (and therefore the weak differential form of strong-f-monotonicity) with a strong-f-monotonicity constant $a = \frac{c}{b^2+c^2} > 0$ since

$$w^t M w = c||w||^2 = \frac{c}{b^2 + c^2}(b^2 + c^2)||w||^2 = a||Mw||^2.$$

Example 5: The differential form of monotonicity does not imply the weak differential form of strong-f-monotonicity.

Consider the function $f(x) = Mx$ with the Jacobian matrix $M = \begin{bmatrix} 0 & -b \\ b & 0 \end{bmatrix}$ and $b \neq 0$.

M is a positive semidefinite matrix. $M^t - aM^tM = \begin{bmatrix} -ab^2 & b \\ -b & -ab^2 \end{bmatrix}$ and since $b \neq 0$, there is no value of the constant $a > 0$ for which $M^t - aM^tM$ is a positive semidefinite matrix, since when $b \neq 0$, for all values of $a > 0$, $M^t - aM^tM$ is negative definite. Therefore, $f(x) = Mx$ is not a strongly-f-monotone function (which in this case coincides with the weak differential form of strong-f-monotonicity). □

To this point, in the symmetric case we have shown the relationship between convexity (monotonicity of f) and the weak differential form of strong-f-monotonicity and for the general case, we have shown that the weak differential form of strong-f-monotonicity of the squared Jacobian matrix (and the positive semidefiniteness of the Jacobian) imply the monotonicity condition. To carry this analysis further, we now consider the relationship between the weak differential form of strong-f-monotonicity and the p.s.d. plus conditions.

In stating the following result, we assume that $\nabla f(x)$ is a p.s.d. plus matrix and therefore from Lemma 3.1 it can be rewritten as $\nabla f(x) = P(x)^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix} P(x)$. We let $Q(x)$ be a submatrix of $P(x)^t P(x)$ defined as follows: let I be an identity matrix with the same dimension as $P_0(x)$; then $\begin{bmatrix} 0 & 0 & 0 \\ 0 & Q(x) & 0 \\ 0 & 0 & 0 \end{bmatrix} = P(x) \begin{bmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{bmatrix} P(x)^t$, so $Q(x) = D(x)D(x)^t + E(x)E(x)^t + F(x)F(x)^t$ when $P(x) = \begin{bmatrix} A(x) & B(x) & C(x) \\ D(x) & E(x) & F(x) \\ G(x) & H(x) & J(x) \end{bmatrix}$.

Proposition 3.12:

Suppose that the matrix $\nabla f(x)$ is p.s.d. plus for all $x \in K$, then the weak differential form of the strong-f-monotonicity condition holds whenever the maximum eigenvalue of the matrix

$$B(x) = [\frac{P_0(x) + P_0(x)^t}{2}]^{-\frac{1}{2}} P_0(x)^t Q(x) P_0(x) [\frac{P_0(x) + P_0(x)^t}{2}]^{-\frac{1}{2}}$$

is bounded over the feasible set K by a constant d.

Conversely, if the weak differential form of the strong-f-monotonicity condition holds, then the matrix $\nabla f(x)$ is a p.s.d. plus.

Proof: Suppose $\nabla f(x)$ is a p.s.d. plus matrix. Then,

$$w^t \nabla f(x) w = w^t P(x)^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix} P(x) w = w^t P(x)^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{P_0(x)+P_0(x)^t}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} P(x) w =$$

$$[v^t, y^t, z^t]\begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{P_0(x)+P_0(x)^t}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} v \\ y \\ z \end{bmatrix},$$

with $P(x)w = \begin{bmatrix} v \\ y \\ z \end{bmatrix}$ and y a vector with the same dimension as $P_0(x)$. Then since $P_0(x)$ is positive definite

$$w^t\nabla f(x)w = y^t[\frac{P_0(x)+P_0(x)^t}{2}]y = y^t[\frac{P_0(x)+P_0(x)^t}{2}]^{\frac{1}{2}}.[\frac{P_0(x)+P_0(x)^t}{2}]^{\frac{1}{2}}y.$$

Furthermore, $w^t\nabla f(x)^t\nabla f(x)w = w^tP(x)^t\begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x)^t & 0 \\ 0 & 0 & 0 \end{bmatrix}P(x)P(x)^t\begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix}P(x)w =$

$$[0, y^tP_0(x)^t, 0]P(x)P(x)^t\begin{bmatrix} 0 \\ P_0(x)y \\ 0 \end{bmatrix} = y^tP_0(x)^tQ(x)P_0(x)y =$$

$$y^t[\frac{P_0(x)+P_0(x)^t}{2}]^{\frac{1}{2}}[\frac{P_0(x)+P_0(x)^t}{2}]^{-\frac{1}{2}}P_0(x)^tQ(x)P_0(x)[\frac{P_0(x)+P_0(x)^t}{2}]^{-\frac{1}{2}}[\frac{P_0(x)+P_0(x)^t}{2}]^{\frac{1}{2}}y.$$

Therefore, if $b = [\frac{P_0(x)+P_0(x)^t}{2}]^{\frac{1}{2}}y$, then $w^t\nabla f(x)w = b^tb$ and $w^t\nabla f(x)^t\nabla f(x)w = b^t[\frac{P_0(x)+P_0(x)^t}{2}]^{-\frac{1}{2}}P_0(x)^tQ(x)P_0(x)[\frac{P_0(x)+P_0(x)^t}{2}]^{-\frac{1}{2}}b.$

Since the maximum eigenvalue of $B(x)$ is bounded over the feasible set K by a constant d, then $\frac{b^tQ(x)b}{b^tb} \le d$ and so for $a = \frac{1}{d}$ we have $w^t\nabla f(x)w \ge aw^t\nabla f(x)^t\nabla f(x)w$ for all $w \in R^n$ and $x \in K$.

Conversely, if for some constant $a > 0$, $w^t\nabla f(x)w \ge aw^t\nabla f(x)^t\nabla f(x)w$ for all $x \in K$ and $w \in R^n$ then, as we have already observed previously, $\nabla f(x)$ is a p.s.d. matrix and therefore p.s.d. plus. □

Remark: In the symmetric case, the matrix $B(x) = [\frac{P_0(x)+P_0(x)^t}{2}]^{-\frac{1}{2}}P_0(x)^tQ(x)P_0(x)[\frac{P_0(x)+P_0(x)^t}{2}]^{-\frac{1}{2}}$ becomes $B(x) = ([P_0(x)]^{\frac{1}{2}})^tQ(x)[P_0(x)]^{\frac{1}{2}}$.

Furthermore, $B(x) = D(x)$ since $P_0(x) = D(x)$ is a diagonal matrix whose diagonal elements are the positive eigenvalues $d_i(x)$ of $\nabla f(x)$ and $Q(x) = I$. Therefore, requiring the maximum eigenvalue of $B(x)$ to be bounded over the feasible set K coincides with the assumption of Proposition 3.9, i.e., $sup_{x \in K}[max_i d_i(x)] \leq d$. So Proposition 3.12 is a natural generalization of Proposition 3.9.

Corollary 3.1: ([21])

Suppose a variational inequality problem is affine with $f(x) = Mx - c$. Then the matrix M is p.s.d. plus if and only if its problem function f is strongly-f-monotone.

The proof of this result follows directly from Proposition 3.12 since in the affine case the weak differential form of strong-f-monotonicity coincides with the regular differential form of strong-f-monotonicity.

Proposition 3.9 shows the relationship between convexity and the weak differential form of strong-f-monotonicity. For the asymmetric case, the analog of convexity is monotonicity. Therefore, we might wish to address the following question. *What is the relationship between monotonicity and strong-f-monotonicity for the general asymmetric case?*

Example 4 in the proof of Proposition 3.11 shows that monotonicity does not imply strong-f-monotonicity. What additional conditions do we need to impose on the feasible set K and the problem function f other than compactness to ensure that monotonicity implies strong-f-monotonicity? Example 3 suggests that even in the symmetric case compactness and convexity are not enough. We need to impose additional assumptions. For this development, we use the following definition which applies to general asymmetric matrices.

Definition 2 : *A matrix $M(x)$ is uniformly positive semidefinite plus (uniformly p.s.d. plus) if for every point $x \in K$, we can express $M(x)$ as* $M(x) = P^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix} P$, *with P independent of x. $P_0(x)$ is a positive definite or zero matrix of fixed dimension $n_1 \times n_1$, and is always in the same location in the bracketed matrix.*

Remark: As our nomenclature shows every uniformly p.s.d. plus matrix $M(x)$ is also p.s.d. plus.

Before continuing, we state a preliminary result about uniformly p.s.d. plus matrices. We first set some notation. In the representation of a p.s.d. plus matrix as specified in Definition 2, suppose we partition P compatibly with $\begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x) & 0 \\ 0 & 0 & 0 \end{bmatrix}$ as $P = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix}$.

Proposition 3.13:

Suppose $\nabla f(x)$ is a uniformly p.s.d. plus matrix. Then f is strongly-f-monotone whenever, for the values of x_1 for which the matrix $P_0(x_1)$ is positive definite, the maximum eigenvalue of the matrix $B(x_1,x_2)^t B(x_1,x_2)$ is bounded over the feasible set K by some constant d^2, with

$$B(x_1,x_2) = [\frac{P_0(x_1)+P_0(x_1)^t}{2}]^{-\frac{1}{2}} P_0(x_1)^t Q P_0(x_2) [\frac{P_0(x_1)+P_0(x_1)^t}{2}]^{-\frac{1}{2}}.$$

Proof: First, we observe that if $x_1 \in K$ and $P_0(x_1) = 0$, then $\nabla f(x_1) = 0$ and, therefore, for all $a > 0$ and $x_2 \in K$, $aw^t \nabla f(x_1)^t \nabla f(x_2) w \leq w^t \nabla f(x_1)^t w$, which is the differential form of strong-f-monotonicity. Now suppose that $x_1 \in K$ and $P_0(x_1)$ is positive definite. The uniform p.s.d. plus property implies that

$$w^t \nabla f(x_1)^t \nabla f(x_2) w = w^t P^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x_1)^t & 0 \\ 0 & 0 & 0 \end{bmatrix} P P^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_0(x_2) & 0 \\ 0 & 0 & 0 \end{bmatrix} P w.$$

Then

$$w^t \nabla f(x_1)^t \nabla f(x_2) w = w^t P^t \begin{bmatrix} 0 & 0 & 0 \\ P_0(x_1)^t D & P_0(x_1)^t E & P_0(x_1)^t F \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & D^t P_0(x_2) & 0 \\ 0 & E^t P_0(x_2) & 0 \\ 0 & F^t P_0(x_2) & 0 \end{bmatrix} P w =$$

$$[v^t, y^t, z^t] \begin{bmatrix} 0 & 0 & 0 \\ P_0(x_1)^t D & P_0(x_1)^t E & P_0(x_1)^t F \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & D^t P_0(x_2) & 0 \\ 0 & E^t P_0(x_2) & 0 \\ 0 & F^t P_0(x_2) & 0 \end{bmatrix} \begin{bmatrix} v \\ y \\ z \end{bmatrix},$$

with $Pw = \begin{bmatrix} v \\ y \\ z \end{bmatrix}$, and y a vector with the same dimension as $P_0(x_i)$, $i = 1, 2$. Let

$Q = EE^t + DD^t + FF^t$.

Therefore, since $\frac{P_0(x_1)+P_0(x_1)^t}{2}$ is a positive definite and symmetric matrix,

$$w^t \nabla f(x_1)^t \nabla f(x_2) w = y^t P_0(x_1)^t Q P_0(x_2) y =$$

$$y^t [\frac{P_0(x_1) + P_0(x_1)^t}{2}]^{\frac{1}{2}} [\frac{P_0(x_1) + P_0(x_1)^t}{2}]^{-\frac{1}{2}} P_0(x_1)^t Q P_0(x_2) [\frac{P_0(x_1) + P_0(x_1)^t}{2}]^{-\frac{1}{2}} [\frac{P_0(x_1) + P_0(x_1)^t}{2}]^{\frac{1}{2}}$$

If $b = [\frac{P_0(x_1)+P_0(x_1)^t}{2}]^{\frac{1}{2}} y$ then $w^t \nabla f(x_1)^t w = b^t b$ and

$$w^t \nabla f(x_1)^t \nabla f(x_2) w = b^t [\frac{P_0(x_1) + P_0(x_1)^t}{2}]^{-\frac{1}{2}} P_0(x_1)^t Q P_0(x_2) [\frac{P_0(x_1) + P_0(x_1)^t}{2}]^{-\frac{1}{2}} b =$$

$$b^t B(x_1, x_2) b \leq b^t b ||B(x_1, x_2)|| \leq d b^t b = d w^t \nabla f(x_1)^t w,$$

since the maximum eigenvalue of $B(x_1, x_2)^t B(x_1, x_2)$ is bounded over the feasible set K by a constant d^2. This inequality shows that for all $x_1, x_2 \in K$ and $w \in R^n$ the constant $a = min\{\frac{1}{d}, 1\} > 0$ satisfies the condition,

$$a w^t \nabla f(x_1)^t \nabla f(x_2) w \leq w^t \nabla f(x_1)^t w,$$

which is the differential form of strong-f-monotonicity (see Table II). Therefore, f is strongly-f-monotone. □

Remarks:

(1) In Proposition 3.12 we have shown that when $\nabla f(x)$ is uniform p.s.d. plus the weak differential form of strong-f-monotonicity holds, that is for some constant $a > 0$ and for all $w \in R^n$ and $x_1 \in K$,

$$w^t \nabla f(x_1) w \geq a w^t \nabla f(x_1)^t \nabla f(x_1) w. \quad (6)$$

The proof of Proposition 3.13 permits us to show that when $\nabla f(x)$ is uniform p.s.d. plus and K is compact, the weak form and usual differential form of strong-f-monotonicity are equivalent. To establish this result, we note that the steps of Proposition 3.13 and the fact that the matrix $[P_0(x_1)^t Q P_0(x_1)]$ is positive definite and symmetric imply the following result. Let $B(x_1, x_2)$ be defined as in Proposition 3.13 and let d^2 be the maximum eigenvalue of $B(x_1, x_2)^t B(x_1, x_2)$ over the compact set K.

$$w^t \nabla f(x_1)^t \nabla f(x_2) w = y^t P_0(x_1)^t Q P_0(x_2) y =$$

$$y^t[P_0(x_1)^tQP_0(x_1)]^{\frac{1}{2}}[P_0(x_1)^tQP_0(x_1)]^{-\frac{1}{2}}P_0(x_1)^tQP_0(x_2)[P_0(x_1)^tQP_0(x_1)]^{\frac{1}{2}}[P_0(x_1)^tQP_0(x_1)]^{\frac{1}{2}}y =$$

$$b^tB(x_1,x_2)b \leq ||B(x_1,x_2)||b^tb \leq db^tb =$$

$$y^tP_0(x_1)^tQP_0(x_1)y = w^t\nabla f(x_1)^t\nabla f(x_1)w$$

with $b = [P_0(x_1)^tQP_0(x_1)]^{\frac{1}{2}}y$. Therefore, $w^t\nabla f(x_1)^t\nabla f(x_2)w \leq dw^t\nabla f(x_1)^t\nabla f(x_1)w$, which implies that the weak and usual form of strong-f-monotonicity are equivalent.

(2) Proposition 3.6 and remark (1) imply that if $\nabla f(x)$ is uniformly p.s.d. plus and K is compact, then for the general iterative scheme (4) (see [7], [25]), the differential form of strong-f-monotonicity is equivalent to the norm condition in a less than or equal form, i.e., $||g_x^{-1/2}(x,x)g_y(x,x)g_x^{-1/2}(x,x)|| \leq 1 \quad \forall x \in K$.

Definition 3 *: (Sun [28]) A matrix $M(x)$ satisfies the Hessian similarity property over the set K if (i) $M(x)$ is a positive semidefinite matrix for all $x \in K$, and (ii) for all $w \in R^n$ and $y, z \in K$ and for some constant $r \geq 1$, $M(x)$ satisfies the condition*

$$rw^tM(z)w \geq w^tM(y)w \geq \frac{1}{r}w^tM(z)w.$$

Matrices that do not depend on x, i.e., $M = M(x)$ for all x, and positive definite matrices on compact sets K satisfy this property. In the later case, we can choose r as the ratio of the maximum eigenvalue of $M(x)$ over K divided by the minimum eigenvalue of $M(x)$ over K.

Sun [28] has established the following result.

Lemma 3.2:

If a matrix is positive semidefinite and symmetric and satisfies the Hessian similarity property then it also satisfies the uniform p.s.d. plus property.

Corollary 3.2:

If for a variational inequality problem, $\nabla f(x)$ is a symmetric, positive definite matrix and the set K is compact (i.e., strictly convex minimization problems on compact sets), then $\nabla f(x)$ satisfies the uniform p.s.d. plus property and the problem function f is strongly-f-monotone.

Proof: When $\nabla f(x)$ is a symmetric, positive definite matrix and the set K is compact, $\nabla f(x)$ satisfies the Hessian similarity condition. Lemma 3.2 implies the uniform p.s.d. plus property. Therefore, Proposition 3.13 implies that f is a strongly-f-monotone problem function. □

Corollary 3.3:

If the Jacobian matrix $\nabla f(x)$ of a variational inequality problem is symmetric and positive semidefinite and satisfies the Hessian similarity condition and the set K is compact, then the problem function f is strongly-f-monotone.

Proof: By Lemma 3.2, the Jacobian matrix $\nabla f(x)$ satisfies the uniform p.s.d. plus condition and so the result follows from Proposition 3.13.

The following result provides a generalization of Proposition 3.13.

Proposition 3.14:

Suppose that $\nabla f(x)$ can be written as

$$\nabla f(x) = P^t \begin{bmatrix} P_1(x) & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & P_i(x) & \dots & 0 \\ \vdots & \dots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & P_m(x) \end{bmatrix} P,$$

the matrices $P_i(x)$ for $i = 1, 2, ..., m$ are either positive definite or zero and for all $i = 1, 2, ..., m$ and they have the same dimension $n_1 \times n_1$ for all x, moreover $PP^t = I$. Let

$$B_i(x_1, x_2) = [\frac{P_i(x_1) + P_i(x_1)^t}{2}]^{-\frac{1}{2}} P_i(x_1)^t Q P_i(x_2) [\frac{P_i(x_1) + P_i(x_1)^t}{2}]^{-\frac{1}{2}},$$

then f is a strongly-f-monotone function, whenever, for $i = 1, ..., m$ and for the values of x_1 for which the matrix $P_i(x_1)$ is positive definite, the matrix $B_i(x_1, x_2)^t B_i(x_1, x_2)$ has maximum eigenvalue that is bounded over the feasible set K by a constant d_i^2.

Proof: We will first define the matrix $D_i(x) = P^t \begin{bmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & P_i(x) & \dots & 0 \\ \vdots & \dots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix} P.$

Then $\nabla f(x) = \sum_{i=1}^{m} D_i(x)$. Observe that since $PP^t = I$, $D_i(x_1)D_j(x_2) = 0$ for $i \neq j$. Therefore, Proposition 3.13 permits us to conclude that

$$aw^t \nabla f(x_1)^t \nabla f(x_2) w = aw^t \sum_{i,j=1}^{m} (D_i(x_1)^t D_j(x_2)) w =$$

$$aw^t \sum_{i=1}^{m} (D_i(x_1)^t D_i(x_2)) w = a \sum_{i=1}^{m} [w^t D_i(x_1)^t D_i(x_2) w] \leq \sum_{i=1}^{m} w^t D_i(x_1)^t w = w^t \nabla f(x_1)^t w$$

for $a = 1/d$ and $d = max_{\{i=1,\dots,m\}} d_i$.

Corollary 3.4:

If the Jacobian matrix $\nabla f(x)$ of a variational inequality problem is a diagonal positive semidefinite matrix and the set K is compact, the problem function f is strongly-f-monotone.

Proof: The proof of this result follows directly from Proposition 3.14, since the diagonal positive semidefinite matrix $\nabla f(x)$ is the sum of uniform p.s.d. plus matrices, with $P_i(x)$ as 1×1 matrices that are zero or positive definite (zero or positive scalars in this case), and with $P = I$.

Remarks:

(i) In Proposition 3.14 we could have made a more general "orthogonality" assumption that $D_i(x_1)D_j(x_2) = 0$ for $i \neq j$ and $\forall x_1, x_2$, which is the central observation in its proof. Then Corollary 3.1, 3.2, 3.3 and 3.4 would become special cases of Proposition 3.14.

(ii) The condition on $\nabla f(x)$ in Proposition 3.14 require that each matrix $P_i(x)$ has fixed dimensions and occupies a fixed location in the block diagonal matrix of the $P_i(x)$s. Can these conditions be relaxed in any sense? Doing so would permit us to define a broader class for which a p.s.d. plus type of condition would imply strong-f-monotonicity.

Finally, we note that strong-f-monotonicity is related to the condition of firmly nonexpansiveness (used, for example, by Lions and Mercier [15], Bertsekas and Eckstein [8]).

Definition 4 *: 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,

Proposition 3.15:

If a problem function f is strongly-f-monotone for a constant $a \geq 1$, then it is firmly nonexpansive. Conversely, if a problem function f is firmly nonexpansive, then it is strongly-f-monotone for the constant $a = 1$.

Remark:

To conclude this discussion, we note that most of the results in this paper, including the orthogonality theorem, can be easily extended to a more general form of a variational inequality,

$$\text{find } x^{opt} \in K : \quad f(x^{opt})^t(x - x^{opt}) + F(x) - F(x^{opt}) \geq 0, \quad \forall x \in K,$$

with $f : K \rightarrow R^n$ a continuous function, $F : K \rightarrow R$ a continuous and convex function, and K a closed and convex subset of R^n.

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] 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.

[3] 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.

[4] G. Cohen, "Auxiliary problem principle extended to variational inequalities", *Journal of Optimization Theory and Applications*, 59, 325-333.

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

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

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

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

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

[16] Z. Luo and P. Tseng (1991), "A decomposition property for a class of square matrices", *Applied Mathematical letters*, 4, 67-69.

[17] T.L. Magnanti and G. Perakis (1993), "A unifying geometric framework for solving variational inequality problems", Working paper OR 276-93, Operations Research Center, M.I.T.

[18] T.L. Magnanti and G. Perakis (1993), "On the convergence of classical variational inequality algorithms", presentation at the ORSA-TIMS national meeting, Chicago, May 1993.

[19] 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.

[20] T.L. Magnanti and G. Perakis (1993), " From Frank-Wolfe to steepest descent; A descent framework for solving *VIPs*", in preparation.

[21] 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.

[22] B. Martos (1975), *Nonlinear programming: theory and methods*, North-Holland, Amsterdam.

[23] A. Nagurney (1992), *Network economics: a variational inequality approach*, Kluwer Academic Publishers, Hingham MA.

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

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

[26] 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.

[27] 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.

[28] J. Sun, (1990), "An affine scaling method for linearly constrained convex programs", (Preprint), Department of Industrial Engineering and Management Sciences, Northwestern University.

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

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

TABLES

What Theorems	Which Algorithms	What Conditions
Banach	Projection [25], [7], [5], [3]	strong monot., $0 < \rho < 2a/L^2 g$
Banach	Relaxation [1], [6]	$\sup_{x,y \in K} \|g_y(x,y)\| \leq \lambda\alpha$, $0 < \lambda < 1$
Banach	Original Steepest Descent [12]	$Df(x)$ p.d., $Df(x)^2$ p.d.
Banach	Cohen's Aux. Probl. Fram. [4]	strong monot., $0 < \rho < 2ab/L^2$
Banach	Forw. and backw. step alg. [10]	strong monot., $0 < \rho < 2a/g$
Baillon	Averages of Steepest Descent [12]	$Df(x)$ p.d., $Df(x)^2$ p.s.d.
Baillon	Averages of Short Step Steepest Descent [20]	strong-f-monot., $0 < \rho \leq 2a$
Baillon	Averages of Constr. Short Step Steepest Descent [20]	strong-f-monot., $0 < \rho \leq 2a$
Baillon	Averages of Projection [19]	strong-f-monot., $0 < \rho \leq 2a/g$
Baillon	Averages of Relaxation [19]	$\sup_{x,y \in K} \|g_y(x,y)\| \leq \alpha$
Opial	Forw.-backw. oper. splitting alg. [10]	strong-f-monot., $0 < \rho < 2a/g$
Opial	Projection [10]	strong-f-monot., $0 < \rho < 2a/g$
Opial	Cohen's Aux. Probl. Framew. [21]	strong-f-monot., $0 < \rho < 2ab$
Opial	Short Step Steepest Descent [20]	strong-f-monot., $0 < \rho < 2a$
Opial	Constr. Short Step Steepest Descent [20]	strong-f-monot., $0 < \rho < 2a/g$
Opial	Asymmetric Projection [29]	G asym., p.d., $G'^{-1/2}[f(G'^{-1/2}y) - (G - G')G'^{-1/2}y]$ strong-f-mon., cnst. $\geq 1/2$
Opial	Modified Aux. Probl. Framew. [21]	$f - M$ strong mon., K' strong mon., some ρ
Orthogonality	Short Step Steepest Descent [20]	strong-f-monot., $0 < \rho < 2a$
Orthogonality	Constr. Short Step Steepest Descent [20]	strong-f-monot., $0 < \rho < 2a/g$
Orthogonality	Projection [19]	strong-f-monot., $0 < \rho < 2a/g$
Orthogonality	Accum. pts. of Sym. Frank-Wolfe [19]	strong-f-monot., symmetry, $K \neq$, comp., conv.
Orthogonality	Accum. pts. of Affine Asym. Frank-Wolfe [20]	affine, strong monot., near-square-symmetry
Orthogonality	Accum. pts. of Affine descent Fram. [20]	affine, strong monot., near-square-symmetry
Orthogonality	Accum. pts. of Geometric Framework [17]	strong-f-monot., $K \neq$, convex, compact
Orthogonality	Cohen's Aux. Probl. Framew.	strong-f-monot., $0 < \rho < 2ab$
Orthogonality	Asymmetric Projection	G asym., p.d., $G'^{-1/2}[f(G'^{-1/2}y) - (G - G')G'^{-1/2}y]$ strong-f-mon, cnst. $\geq 1/2$
Orthogonality	Modified Aux. Probl. Framew.	$f - M$ strong mon., K' strong mon., some ρ

Table I. Convergence approaches

Notes:

In this table, G' denotes the symmetric part of the matrix G involved in the projection method, i.e., $G' = \frac{1}{2}(G + G^t)$. The constants involved are $g = min(eigenvalue\ of\ G)$, a is the strong-f-monotonicity constant, L is the Lipschitz continuity constant, $\alpha = \inf_{x,y \in K}\ (min(eigenvalue\ g_x(x,y))) > 0$, b is the constant involved in the strong convexity of the function K. Finally, the map M is the monotone part of the function G_ϵ that is involved in the modified auxiliary problem framework (see [21] for more details).

Type of monotonicity imposed upon f	Definition*	Differential condition*
monotone on K	$[f(x) - f(y)](x - y) \geq 0$	$\nabla f(x)$ p.s.d.$^{+}$
strongly-f-monotone on K	$\exists a > 0,\ [f(x) - f(y)](x - y) \geq a \parallel f(x) - f(y) \parallel_2^2$	$\exists a > 0,\ [\nabla f(x)^t - a \nabla f(x)^t \nabla f(y)]$ p.s.d.$^{+}$
strictly strongly-f-monotone on K^{***}	$\exists a > 0,\ [f(x) - f(y)](x - y) > a \parallel f(x) - f(y) \parallel_2^2$	$\exists a > 0,\ [\nabla f(x)^t - a \nabla f(x)^t \nabla f(y)]$ p.d.$^{++}$
strictly monotone on K^{**}	$[f(x) - f(y)](x - y) > 0$	$\nabla f(x)$ p.d.$^{++}$
strongly monotone on K^{**}	$\exists a > 0,\ [f(x) - f(y)](x - y) \geq a \parallel x - y \parallel_2^2$	$\nabla f(x)$ uniformly p.d.$^{++}$
* Definition holds for all $x, y \in K$ or all $x \in K$ ** Condition holds for $x \neq y$ *** Condition holds for $f(x) \neq f(y)$		+ p.s.d. means positive semidefinite ++ p.d. means positive definite

Table II. Several types of monotonicity

FIGURE I

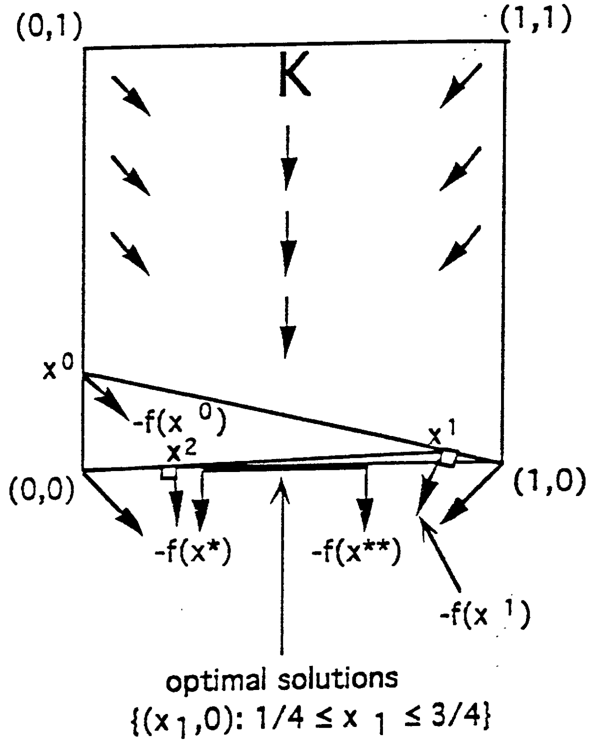
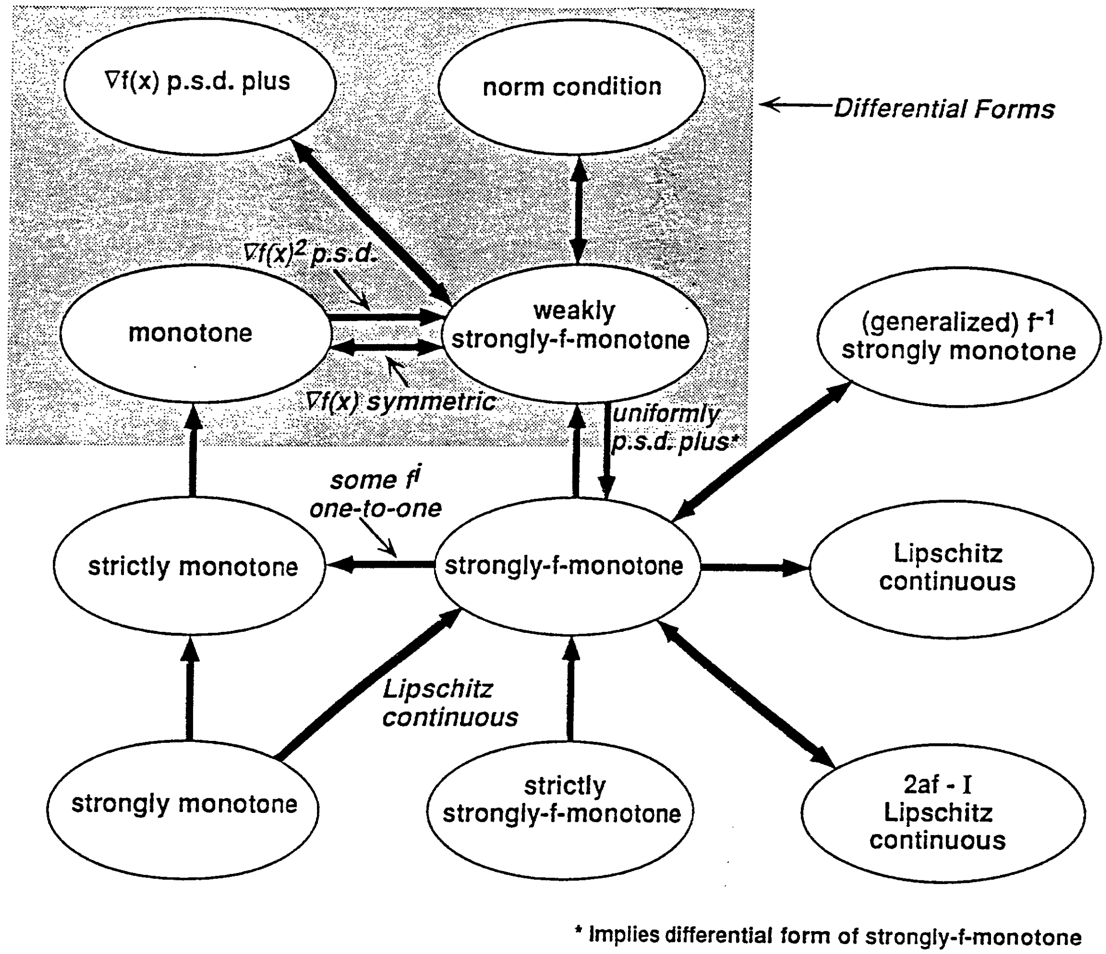


Figure 1: The orthogonality theorem.

FIGURE II.



Relationships between different types of monotonicity