On the Convergence of Classical Variational Inequality Algorithms

Thomas L. Magnanti *Georgia Perakis*

OR 280-93 May 1993

On the Convergence of Classical Variational Inequality Algorithms

Thomas L. Magnanti * Georgia Perakis †‡

May 1993

Abstract

In this paper, we establish global convergence results for projection and relaxation algorithms for solving variational inequality problems, and for the Frank-Wolfe algorithm for solving convex optimization problems defined over general convex sets. The analysis rests upon the condition of f-monotonicity, which we introduced in a previous paper, and which is weaker than the traditional strong monotonicity condition. As part of our development, we provide a new interpretation of a norm condition typically used for establishing convergence of linearization schemes. Applications of our results arize in uncongested as well as congested transportation networks.

Dedication

We dedicate this paper to Stella Dafermos, who was not only instrumental in launching this research, but also a dear friend and colleague.

---

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

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

‡The second author wishes to acknowledge the Applied Mathematics Department at Brown University. Part of this work was completed while she was a student there.

1 Introduction

In this paper we consider the variational inequality problem

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

defined over a compact, 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.

Let $MIN(F,K)$ denote an optimization problem with an objective function $F(x)$ and constraints $x \in K$. The function $f(x)$ of the variational inequality problem $VI(f,K)$ associated with the minimization problem $MIN(F,K)$ is defined by $f(x) = \nabla F(x)$. As is well-known, a variational inequality problem $VI(f,K)$ is equivalent to a (strictly) convex nonlinear programming problem $MIN(F,K)$ (that is they have the same set of solutions) if and only if the Jacobian matrix of the problem function f is a symmetric and positive semidefinite (definite) matrix.

The literature contains a substantial number of algorithms for the numerical solution of the variational inequality problem. The Ph.D. thesis of Hammond [16] and the review paper of Harker and Pang [19] summarize and categorize many algorithms for the problem.

Projection and relaxation algorithms have been among the most popular, classical algorithms for solving the variational inequality problem (1). Goldstein [15], and independently by Levitin and Polyak [34], first proposed projection algorithms in the context of the nonlinear programming problem. Several authors, including Sibony [35], Bakusinskii and Polyak [5], Auslender [2] and Dafermos [12], have studied projection algorithms for variational inequalities while Dafermos [10], Bertsekas and Gafni [6] and others have studies these algorithms for the traffic equilibrium problem. The projection algorithms can be viewed as special cases of the linearization algorithms developed by Chan and Pang [32] (for an affine problem function f see also the contracting ellipsoid algorithm introduced by Hammond and Magnanti [18]). 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 [12], [11]. All these algorithms are special cases of a general iterative framework developed by Dafermos [12]. Researchers have established the convergence of projection algorithms under the condition of strong monotonicity, and the convergence of the linearization algorithms, and the generalized contracting ellipsoid methods under a norm condition implying strong monotonicity. Relaxation algorithms converge under a stronger norm condition that also implies strong monotonicity. All these methods generate a sequence of points $\{x_k\}_{k=0}^{\infty}$ in the feasible set K, whose convergence to an optimal solution follows from a contraction estimate of the form:

$$||x_{k+1} - x_k||_G \leq a||x_k - x_{k-1}||_G, \quad 0 < a < 1.$$

In this expression, $||.||_G$ denotes the fixed norm in R^n induced by a symmetric, positive definite matrix G as $||x||_G = (x^t G x)^{1/2}$.

Another classical algorithm for solving convex programs is the Frank-Wolfe algorithm. The Frank-Wolfe algorithm, when applied to convex programming problems defined over polyhedra, is a linear approximation method that iteratively approximates the objective function $F(x)$ by $F^k(x) = F(x^k) + \nabla F(x^k)^t(x - x^k)$. On the (k+1)th iteration, the algorithm determines a vertex solution y^{k+1} to the linear program

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

and then chooses as the next iterate the point x^{k+1}, that minimizes the objective function F over the line segment $[x^k; y^{k+1}]$. Frank and Wolfe [13] originally proposed this algorithm for solving quadratic programming problems. The 1975 book of Martos [28] and the Ph.D. thesis of Hammond [16] illustrate the method's performance on several examples and discuss its convergence properties. This algorithm is particularly effective for solving large-scale traffic equilibrium problems; see [8], [21], [14] for further details. In the context of the traffic equilibrium problem, the linear programming component of the algorithm decomposes into a set of shortest path problems, one for each origin-destination pair. Therefore, the Frank-Wolfe algorithm applied in the traffic equilibrium example, first solves a set of shortest path problems and then a one-dimensional minimization problem. The algorithm is known to converge if the objective function F is pseudoconvex and the feasible set K is a bounded polyhedron (see Martos [28]).

Our goal in this paper is to study the convergence behavior of these classical algorithms. Our principal objective is to establish their convergence under assumptions on the given problem function f that are weaker than the existing ones (strong monotonicity, the norm condition) or that extend the applicability of the algorithm (to general convex sets for the Frank-Wolfe algorithm). We also want to establish global convergence proofs for these algorithms that do not depend on the neighborhood in which we initiate the algorithms. In our convergence proofs we obtain nonexpansive estimates (rather than contractive one) of the form

$$||x_{k+1} - x_k||_G \leq ||x_k - x_{k-1}||_G.$$

We use an f-monotonicity condition on the problem function f, that we introduced in [26] and [33]. We say that the problem function f is f-monotone if there exists a positive constant $a > 0$ such that

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

Finally, we want to provide a better understanding of existing conditions such as Chan and Pang's norm condition, and show that there is some "equivalency" between this norm condition (imposed on the algorithm function) and the f-monotonicity condition imposed on the problem function.

This paper is organized as follows. In Section 2 we review some properties of the f-monotonicity condition. We also show the relationship between this condition and the norm condition used by Chan and Pang for the convergence of the linearization algorithms [32] and by Hammond and Magnanti for the convergence of the generalized contracting ellipsoid algorithm [18]. In Section 3, we establish the convergence of the sequence of averages induced by the projection algorithm for the VIP (1), under the condition of f-monotonicity. To achieve this result, we use an ergodic theorem for nonlinear nonexpansive maps due to Baillon [3]. Furthermore, we show that every accumulation point of the projection algorithm sequence solves the variational inequality problem, without a boundedness assumption imposed on the feasible set K. In Section 4 we demonstrate the convergence of the sequence of averages induced by the relaxation scheme for the VIP (1), under a suitable norm condition. We again use Baillon's ergodic theorem. The results in Section 3 and 4 depart from the literature in two ways: (i) we obtain global convergence results, and (ii) we impose weaker assumptions than those required in prior results. For relaxation algorithms, our convergence results are weaker than those in the literature, however, since they apply to the sequence of averages rather than for the sequence itself. In Section 5 we establish the convergence of the Frank-Wolfe algorithm for convex optimization problems over arbitrary compact, convex sets, when the gradient of the objective function satisfies the f-monotonicity condition. To the best of our knowledge, this is the first convergence proof for the Frank-Wolfe algorithm for general convex sets (instead of polyhedra). In Section 6 we examine applications of these results for equilibrium problems in congested as well as uncongested transportation networks.

One motivation for this research is the desire to model and solve large transportation networks, in which some links are uncongested. Our condition of f-monotonicity gives us this possibility, while strict and strong monotonicity do not. Moreover, we show that if we impose the f-monotonicity condition on the link cost function then this condition applies to the path cost function as well. Therefore, the projection algorithm can be applied directly in the space of path flows which is advantageous because it is much easier to perform projection operations efficiently in the space of path flows than in the space of link flows. Another main advantage of our development is that linear programs, when modeled as variational inequalities, have f-monotone problem functions. Therefore, our results apply to linear programs.

Finally, in Section 7, 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 $(x,y)_S = x^tSy$. The inner product induces a norm with respect to the matrix S via*

$$||x||_S^2 = x^tSx.$$

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

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

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

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

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

So,

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

and, similarly,

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

The *argmin* of a function F over a set K is defined as

$$argmin_{x \in K} F(x) = y \in \{x^* : \quad F(x^*) = min_{x \in K} F(x)\}.$$

2 On the condition of f-monotonicity

2.1 Monotone functions

In the past, researchers have found several different forms of monotonicity to be useful in developing underlying theory and analyzing algorithms for variational inequality problems. Table I shows five different definitions of monotonicity and a differential condition for each situation (when the problem function f is differentiable). Whenever the function satisfies any one of these definitions and f is differentiable, f satisfies the corresponding differential condition. When the set K is open and f is differentiable, each definition is equivalent to the corresponding differential condition. For monotone, strictly monotone, and strongly monotone functions, these results are standard (for example, see [31]); for f-monotone and strictly f-monotone functions, the results are due to Magnanti and Perakis [26].

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.+
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 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 I, Several types of monotonicity

Clearly, any f-monotone function is monotone and any strictly f-monotone function is strictly monotone. As shown by Magnanti and Perakis [26], (i) if f is one-to-one, then f-monotonicity implies strict monotonicity, and (ii) if f is Lipschitz continuous, then strong monotonicity implies f-monotonicity. Therefore, for the class of Lipschitz continuous functions, the class of f-monotone functions lies between the classes of monotone and strongly monotone functions.

Note any constant function, i.e., $f(x) = c$ for all x, is an f-monotone function; any variational inequality with a constant problem function and a polyhedron as the feasible set K is equivalent to a linear program in the sense that x^{opt} solves the variational inequality if and only if x^{opt} solves the linear program $min\{c^t x : x \in K\}$. Therefore, one of the principal attractions of f-monotone functions is the fact that the class variational inequalities with f-monotone functions contains all linear programs. Recall that linear programs do not always have optimal solutions (since the defining polyhedron might be unbounded), and so variational inequalities with f-monotone functions need not have a solution.

Note that for affine functions f (i.e., $f(x) = Mx - c$ for some matrix M and vector c), the differentiable f-monotonicity condition holds if we can

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

As a last preliminary observation about f-monotone functions, we note that f-monotonicity is equivalent to strong monotonicity of the generalized inverse f^{-1} of f in the following sense.

Definition 2 . *The generalized inverse f^{-1} of a problem function f is the point to set map*

$$f^{-1} : f(K) \subseteq R^n \rightarrow 2^K,$$

defined by $f^{-1}(X) = \{x \in R^n : \ f(x) = X\}$.

Definition 3 . *A point to set map $g : R^n \rightarrow 2^{R^n}$ is strongly monotone if for every $x \in g^{-1}(X)$ and $y \in g^{-1}(Y)$*

$$(x - y)^t(X - Y) \geq a||X - Y||_2^2,$$

for some constant $a > 0$.

Proposition 1:

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

The proof of this proposition can be found in [26].

2.2 The f-monotonicity condition implies the norm condition.

In this section we show the relationship between the condition of f-monotonicity and the norm condition used by Chan and Pang for establishing the convergence of the linearization algorithms [32] (see also Hammond and Magnanti's discussion of the convergence of the generalized contracting ellipsoid algorithm [18]). These algorithms fit in the framework of the general iterative scheme developed by Dafermos [12], which works as follows:

STEP 0:

Start with some initial point $x_0 \in K$.

STEP $k+1$:

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

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

We make the following assumptions 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, when evaluated at the point $y = x$, which we denote throughout this chapter by $g_x(x,x)$, is a positive definite and symmetric matrix.

For linearization algorithms, $g(x,y) = \rho f(y) + A(y)(x-y)$ for some positive definite matrix $A(y)$ and constant ρ satisfying $0 < \rho \leq 1$. For the generalized contracting ellipsoid algorithm,

$$g(x,y) = f(y) + \rho(\nabla f(y) + \nabla f(y)^t)(x-y),$$

with $\nabla f(y)$ positive definite and a constant ρ satisfying $0 < \rho \leq 1$. In most cases, ρ is chosen equal to one.

In the context of these algorithms, the norm condition we have been referring to is

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

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

$$g(x,y) = \rho f(y) + A(y)(x-y),$$

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

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

In the generalized contracting ellipsoid algorithm, $A(y) = \nabla f(y) + \nabla f(y)^t$ and $\rho = 1$, so 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 and of Hammond and Magnanti as special cases. This condition is more difficult to verify, however, since it involves different points $x_1, y_1, x_2, y_2, x_3, y_3$.

Our goal in this section is to investigate the relationship between the norm condition and the f-monotonicity condition. The main theorem of this section shows that the differential form of f-monotonicity of f implies the norm condition 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 the statement. Namely, the norm condition implies a weaker form of the differential condition of f-monotonicity.

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

LEMMA 1:

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:

Let $x \in R^n$ 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, $x^tG^{-1/2}AG^{-1/2}x \geq 0 \quad \forall x \in R^n$ and so $G^{-1/2}AG^{-1/2}$ is a positive semidefinite matrix. Q.E.D.

LEMMA 2:

Suppose that the matrix

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

is positive semidefinite for some constant $a > 0$. Let G be a positive definite matrix, g be the minimum eigenvalue of G (a positive definite matrix), and $a_1 \leq ag$. Then $(G^{-1/2} \nabla f(x)G^{-1/2})^t(I - a_1G^{-1/2} \nabla f(x)G^{-1/2})$ is also positive semidefinite.

Proof:

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

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

and replacing $a_1 \leq ag$, and $z = G^{-1/2}y$, $v^tG^{-1}v \leq gv^tv$ we obtain:

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

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

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

LEMMA 3:

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,

$$||I - aB|| \leq 1$$

$$\Leftrightarrow \sup_{y \neq 0} \frac{y^t[I - (aB^t + aB) + (aB)^t(aB)]y}{y^ty} \leq 1$$

$$\Leftrightarrow y^t[I - (aB^t + aB) + (aB)^t(aB)]y \leq y^ty \quad \forall y \in R^n$$

$$\Leftrightarrow 2ay^tBy \geq a^2y^tB^tBy \quad \forall y \in R^n$$

$$\Leftrightarrow y^tBy \geq (a/2)y^tB^tBy \quad \forall y \in R^n \tag{2}$$

$$\Leftrightarrow y^tB^t[I - (a/2)B]y \geq 0 \quad \forall y \in R^n$$

These relationships show that $||I - aB|| \leq 1$ if and only if the matrix $B^t[I - (a/2)B]$ is positive semidefinite. Moreover, (2) 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 holds with $B^t[I - (a/2)B]$ positive definite and $||I - aB|| < 1$. We are now ready to prove the central theorem of this section.

THEOREM 1:

Consider the general linearization scheme for some constant $\rho > 0$, $g(x, x) = \rho f(x)$ and that $g_x(x, x)$ is a positive definite and symmetric matrix. Then the following results are valid.

1. If the differential form of f-monotonicity condition holds, 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 $a > 0$, then the norm condition holds in a less than or equal to form. Namely,

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

2. 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 for some constant $a > 0$, the matrix

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

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

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

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

if we let $G = g_x(x, x)$, the norm condition becomes:
$||G^{-1/2}[\rho \bigtriangledown f(x) - G]G^{-1/2}|| = ||I - \rho G^{-1/2} \bigtriangledown f(x) G^{-1/2}|| \leq 1$.
By assumption, G is a positive definite and symmetric matrix. Let

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

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

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

is positive semidefinite. Lemma 3 implies that for any choice of

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

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

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

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

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

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

is positive semidefinite for some $a > 0$ and $\forall x \in K$.
Let $G = g_x(x, x)$. Since

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

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

setting, as before, $B = G^{-1/2} \bigtriangledown f(x) G^{-1/2}$, we see from Lemma 3 that if

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

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

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

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

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

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

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

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

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

is positive semidefinite. Q.E.D.

Remark:

The differential condition of f-monotonicity implies that the norm condition holds in a less than or equal form. The existing convergence proofs require a strictly inequality form of the norm condition. This happens in our case, when the differential form of f-monotonicity holds in some form of a strict inequality, i.e.,

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

is positive semidefinite and the matrix $\bigtriangledown f(x)$ is nonsingular. The norm condition

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

then holds as a strict inequality. The following Proposition formalizes this result.

Proposition 2:

Consider the general linearization scheme for some constant $\rho > 0$. Suppose that $g(x,x) = \rho f(x)$ and that $g_x(x,x)$ is a positive definite and symmetric matrix. Then the following results are valid.

1. If the differential form of 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 holds as a strict inequality. Namely

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

2. If the norm condition

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

holds, then for some constant $a > 0$, the matrix

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

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

The proof of this Proposition follows from that of Theorem 2 using strict inequalities in place of the less than or equal to inequalities.

Remark:

In particular, for the linearization algorithms $g(x,y) = \rho f(y) + A(y)(x-y)$ and for the generalized contracting ellipsoid algorithm with $A(y) = \nabla f(y) + \nabla f(y)^t$, the norm condition becomes:

$||g_x^{-1/2}(x,x) g_y(x,x) g_x^{-1/2}(x,x)|| = ||A(x)^{-1/2}[\rho \nabla f(x) - A(x)] A(x)^{-1/2}|| < 1$
$\forall x \in K$.

Therefore, these two classes of algorithms converge under the assumption that the problem function f is strictly f-monotone. The advantage of stating the result in this way is that we impose a condition directly on the problem function f and we do not involve the algorithm function g.

The convergence proofs of the linearization and the contracting ellipsoid algorithms ([17], [16], [32]) require that the norm is a strict inequality because they require that the algorithm needs to start with an initial point x_0 that lies close to the VIP solution. In fact, if x^{opt} is a solution to the variational inequality and $c = ||A(x)^{-1/2}[\rho \bigtriangledown f(x) - A(x)]A(x)^{-1/2}|| < 1$, they require that

$$||x_0 - x^{opt}|| < \frac{1-c}{C},$$

for some positive constant C. The closer c is to one, the closer the initial point needs to be to the optimal solution.

From this discussion, we conclude that the differential form of f-monotonicity and the assumption that $\bigtriangledown f(x)$ is positive definite imply that the generalized contracting ellipsoid algorithm and the linearization algorithms converge to an optimal solution of the variational inequality problem.

3 On the convergence of the projection algorithm under f-monotonicity

In this section we show that the f-monotonicity condition implies the convergence of the sequence of averages induced by the projection algorithm. To establish this result we employ ideas from the theory of variational inequalities as well as an ergodic theorem due to Baillon [3]. Furthermore, when the feasible set is compact, we show that that f-monotonicity implies that every accumulation point of the projection algorithm sequence solves the variational inequality. We first recall the classical 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:

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

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 [12] and the linearization algorithms [32], with $g(x,y) = \rho f(y) + A(y)(x-y)$ and $A(y) = G$.

In view of the symmetry and positive definiteness of the matrix $g_x(x,y) = G$, the line integral $\int g(x,y)dx$ defines a function $F : K \times K \rightarrow R^n$ satisfying the property that, for fixed $y \in K$, $F(.,y)$ is strictly convex and $g(x,y) = F_x(x,y)$. Therefore, step $k+1$ of the projection algorithm is equivalent to the strictly convex mathematical programming problem:

$$min_{x \in K} \; F(x, x_k).$$

Consequently, the problem defined in step $k+1$ has a unique solution that can be computed by any appropriate nonlinear programming algorithm. For some fixed $x_k \in K$ (found in the previous step), if we let $h_k = \rho f(x_k) - Gx_k$ we can rewrite $g(x, x_k)$ as $g(x, x_k) = Gx + h_k$. In other words, according to the projection algorithm, at each step $k+1$, we need to determine the point $x_{k+1} \in K$ that satisfies the variational inequality for g, visualized as:

$$(Gx_{k+1} + h_k)^t (x - x_{k+1}) \geq 0, \quad \forall x \in K.$$

In view of the symmetry and positive definiteness of the matrix G, this variational inequality has a unique solution x_{k+1}, which as shown above, is the unique minimum over K, of the function $F(., x_k)$, with x_k fixed from the previous step. For notational convenience, let us define $F_k(x)$ as $F(x, x_k) = F_k(x) = (1/2)x^t G x + h_k^t x$.

3.1 On the convergence of the sequence of averages induced by the projection algorithm

In this section we establish the convergence of the sequence of averages induced by the projection algorithm. The key theorem we employ to establish this result is an ergodic theorem due to Baillon. Throughout this analysis, we assume the feasible set to be compact; this is not an essential assumption. We can derive similar results by just assuming that the feasible set is closed and convex.

For the subsequent analysis, let us define $T_\rho : K \longrightarrow R^n$ as the map that carries $x_k \in K$ into the minimizer over K of the function $F_k(.)$. Namely, $x_{k+1} = T_\rho(x_k)$. The following lemma describes the relevance of the map T_ρ.

LEMMA 4:

Every fixed point of the map T_ρ is a solution of the original asymmetric variational inequality problem (1).

Proof:

The definition of map T_ρ implies that if $x_k = x_{k+1} = T_\rho(x_k) \in K$ is a fixed point of T_ρ, then

$$(Gx_{k+1} + h_k)^t(x - x_{k+1}) \geq 0 \quad \forall x \in K.$$

In this case $h_k = \rho f(x_k) - Gx_k = \rho f(x_{k+1}) - Gx_{k+1}$. Making this replacement in the inequality gives

$$(Gx_{k+1} + \rho f(x_{k+1}) - Gx_{k+1})^t(x - x_{k+1}) = \rho f(x_{k+1})^t(x - x_{k+1}) \geq 0 \quad \forall x \in K.$$

Therefore, $x_{k+1} = T_\rho(x_k) = x_k$ is a solution of the original asymmetric variational inequality problem (1). Q.E.D.

LEMMA 5:

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

$$(Gx_1 + h_1)^t(x - x_1) = (Gx_1 + \rho f(y_1) - Gy_1)^t(x - x_1) \geq 0 \quad \forall x \in K, \qquad (3)$$

$$(Gx_2 + h_2)^t(x - x_2) = (Gx_2 + \rho f(y_2) - Gy_2)^t(x - x_2) \geq 0 \quad \forall x \in K. \qquad (4)$$

Setting $x = x_2$ in (3) and $x = x_1$ in (4) 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

$$||x_1 - x_2||_G^2 \leq ||y_1 - y_2||_G^2 - 2\rho[f(y_1) - f(y_2)]^t[y_1 - y_2] +$$
$$(\rho)^2[f(y_1) - f(y_2)]^t G^{-1}[f(y_1) - f(y_2)]. \qquad (5)$$

The f-monotonicity of map f and the symmetry and positive definiteness of matrix G, together with this result, implies that if λ is the minimum eigenvalue of the positive definite, symmetric matrix G, and $b = \frac{1}{\lambda}$, then

$$||x_1 - x_2||_G^2 \leq ||y_1 - y_2||_G^2 - 2\rho a||f(y_1) - f(y_2)||_2^2 + (\rho)^2 b||f(y_1) - f(y_2)||_2^2 =$$

$$= ||y_1 - y_2||_G^2 - \rho(2a - \rho b)||f(y_1) - f(y_2)||_2^2.$$

Finally, since $0 < \rho \leq \frac{2a}{b}$ and $||f(y_1) - f(y_2)||^2 \geq 0$, we conclude that

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

Q.E.D.

Using these lemmas we now establish the convergence of the sequence of averages induced by the projection algorithm. We will use the following ergodic theorem.

THEOREM 2 (Baillon [3]):

Let T be a map, $T : K \longrightarrow K$, 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, that is,

$$||T(y_1) - T(y_2)||_G^2 \leq ||y_1 - y_2||_G^2 \quad \forall y_1, y_2 \in K,$$

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.

In the following theorem we use the finite dimensional version of this theorem.

THEOREM 3:

Let K be a convex, compact subset of R^n (the feasible set of the VIP (1)) and T_ρ : $K \longrightarrow R^n$ be the map that carries $x \in K$ into the minimizer over K of the function $F_k(.)$. Also, let λ be the minimum eigenvalue of the positive definite, symmetric matrix G and let $b = \frac{1}{\lambda}$. Assume in the projection algorithm that $0 < \rho \leq \frac{2a}{b}$. Then if f is a f-monotone map, 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 5, if $0 < \rho \leq \frac{2a}{b}$ in the projection algorithm and if f is f-monotone, then the map T_ρ is nonexpansive relative to the $||.||_G$ norm. Then the finite dimensional version of Theorem 3 guarantees that

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

converges to the fixed point of the map T_ρ. Lemma 4 shows that every fixed point of the map T_ρ is a solution of the original asymmetric variational inequality problem. 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 just a nonexpansive map, but also a contraction map. In this case, the original sequence induced by the projection algorithm converges to the the solution of the VIP (1), which is then unique (since f is strictly monotone). The convergence of the original sequence follows from Banach's fixed point theorem, (see also [10], [12]).

2. Throughout this analysis we have assumed that the feasible set K is compact. We believe that this assumption is not essential. If we start the algorithm with $x_1 \in K$ so that $||T(x_1) - x_1|| \leq \infty$ we lie without loss of generality in a compact set. In the next Section 3.2, we derive similar results for the sequence induced by the projection algorithm, by assuming that the feasible set K is just a closed and convex set.

3.2 Convergence of the projection algorithm

In this section we establish the convergence of the sequence induced by the projection algorithm when the underlying problem function satisfies the f-monotonicity condition. In this proof we will assume that the feasible set K is a closed and convex set and that the VIP problem has at least one optimal solution. We show that every accumulation point x^* of the projection algorithm sequence solves the variational inequality problem. We first prove several lemmas.

LEMMA 6:

Let λ be the minimum eigenvalue of the positive definite and symmetric matrix G and let $b = \frac{1}{\lambda}$. Assume that the variational inequality problem $VI(f, K)$ has at least one optimal solution x^{opt} and that the feasible set K is a closed and convex set. If $0 < \rho < \frac{2a}{b}$ and f is a f-monotone map, then the sequence $\{||x_k - x^{opt}||_G\}_{k=1}^{\infty}$ is a convergent sequence.

Proof:

Step $k + 1$ of the projection algorithm implies that $x_{k+1} \in K$ satisfies the following inequalities

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

Setting $x = x^{opt} \in K$ in Step $k + 1$ and recalling that $[\rho f(x^{opt})]^t[x^{opt} - x_{k+1}] \leq 0$, we see that

$$[\rho f(x_k) - \rho f(x^{opt}) + G(x_{k+1} - x_k)]^t[x^{opt} - x_{k+1}] \geq 0.$$

This inequality implies (by adding $||x_{k+1} - x^{opt}||_G^2$ to both sides) that

$$||x_{k+1} - x^{opt}||_G^2 \leq (x^{opt} - x_k - \rho G^{-1}[f(x^{opt}) - f(x_k)])^t G(x^{opt} - x_{k+1})$$

and Cauchy's inequality implies that

$$||x_{k+1} - x^{opt}||_G^2 \leq ||x^{opt} - x_k - \rho G^{-1}[f(x^{opt}) - f(x_k)]||_G ||x^{opt} - x_{k+1}||_G.$$

Dividing through by $||x^{opt} - x_{k+1}||_G$, squaring, and expanding the righthand side, we obtain:

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

The f-monotonicity of function f and the symmetry and positive definiteness of matrix G implies that

$$||x_{k+1} - x^{opt}||_G^2 \leq ||x_k - x^{opt}||_G^2 - \rho(2a - \rho b)||f(x^{opt}) - f(x_k)||_2^2.$$

Therefore, by choosing $0 < \rho < \frac{2a}{b}$ and by observing that $||f(x^{opt}) - f(x_k)||_2^2 \geq 0$, we see that

$$0 \leq ||x_{k+1} - x^{opt}||_G^2 \leq ||x_k - x^{opt}||_G^2.$$

Therefore, the sequence $\{||x_k - x^{opt}||_G\}_{k=1}^{\infty}$ is a convergent sequence (since it is nonincreasing and bounded from below). Q.E.D.

LEMMA 7:

Let x^{opt} be a solution of the VIP. Under the assumptions of Lemma 6, the sequence $\{f(x_k)\}_{k=1}^{\infty}$ converges to the optimal value $f(x^{opt})$ of the problem.

Proof:

In the course of proof of Lemma 6 we showed that

$$||x_{k+1} - x^{opt}||_G^2 \leq ||x_k - x^{opt}||_G^2 - \rho(2a - \rho b)||f(x^{opt}) - f(x_k)||_2^2.$$

Therefore,

$$||x_k - x^{opt}||_G^2 - ||x_{k+1} - x^{opt}||_G^2 \geq \rho(2a - \rho b)||f(x^{opt}) - f(x_k)||_2^2 \geq 0.$$

Since $||x_k - x^{opt}||_G^2 - ||x_{k+1} - x^{opt}||_G^2 \longrightarrow_{k\to\infty} 0$, and since $0 < \rho < \frac{2a}{b}$, this inequality implies that

$$||f(x^{opt}) - f(x_k)||_2^2 \longrightarrow_{k\to\infty} 0.$$

Thus,

$$f(x_k) \longrightarrow_{k\to\infty} f(x^{opt}).$$

Q.E.D.

LEMMA 8:

Under the assumptions of Lemma 6, the sequence $\{f(x^{opt})^t(x_k - x^{opt})\}_{k=1}^{\infty}$ converges to value zero.

Proof:

Setting $x = x^{opt}$ in Step $k+1$ and recalling that $[\rho f(x^{opt})]^t[x^{opt} - x_{k+1}] \leq 0$, we see that

$$[\rho f(x_k) - \rho f(x^{opt}) + G(x_{k+1} - x_k)]^t[x^{opt} - x_{k+1}] \geq \rho f(x^{opt})]^t[x_{k+1} - x^{opt}] \geq 0.$$

Adding and subtracting $||x_{k+1} - x^{opt}||_G^2$ from the middle term in this expression and rearranging as in Lemma 6 gives

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

Applying Cauchy's inequality, using the fact that f is f-monotone, and following the steps of Lemma 6, we obtain

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

$$\sqrt{||x_k - x^{opt}||_G^2 - \rho(2a - \rho b)||f(x^{opt}) - f(x_k)||_2^2}||x_{k+1} - x^{opt}||_G - ||x_{k+1} - x^{opt}||_G^2.$$

Since $0 < \rho < \frac{2a}{b}$,

$$0 \leq \rho f(x^{opt})]^t[x_{k+1} - x^{opt}] \leq [||x_k - x^{opt}||_G - ||x_{k+1} - x^{opt}||_G][||x_{k+1} - x^{opt}||_G] \longrightarrow_{k \to \infty} 0,$$

due to Lemma 6 (i.e., that the sequence $||x_k - x^{opt}||$ is convergent). Therefore, the sequence $\{f(x^{opt})^t(x_k - x^{opt})\}_{k=1}^{\infty}$ converges to value zero. Q.E.D.

LEMMA 9:

Under the assumptions of Lemma 6, the sequence $\{f(x_k)^t(x_k - x^{opt})\}_{k=1}^{\infty}$ converges to value zero.

Proof:

$$0 \leq f(x_k)^t(x_k - x^{opt}) = [f(x_k) - f(x^{opt}]^t(x_k - x^{opt}) + f(x^{opt})^t(x_k - x^{opt}) \leq$$

(applying Cauchy's inequality)

$$||f(x_k) - f(x^{opt}||_2.||x_k - x^{opt}||_2 + f(x^{opt})^t(x_k - x^{opt}) \longrightarrow_{k \to \infty} 0,$$

due to Lemmas 6,7 and 8. Q.E.D.

THEOREM 4:

If the variational inequality problem $VI(f, K)$ has at least one solution x^{opt}, the feasible set K is a closed, convex subset of R^n, the problem function f is f-monotone, and $0 < \rho < \frac{2a}{b}$ in the projection algorithm, then

$$\lim_{k \to \infty} f(x_k)^t(x - x_k) \geq 0 \ \ \forall x \in K,$$

and every accumulation point of the algorithm solves the variational inequality problem.

Proof:

We intend to show that, under the assumptions of this theorem, $\lim_{k \to \infty} f(x_k)^t(x - x_k)$ exists and,

$$\lim_{k \to \infty} f(x_k)^t(x - x_k) \geq 0 \quad \forall x \in K.$$

Let x^{opt} be an optimal solution of (1). Then if we add and subtract x^{opt} from the lefthand side, we see that $\forall x \in K$:

$$\lim_{k \to \infty} f(x_k)^t(x - x_k) = \lim_{k \to \infty} f(x_k)^t(x - x^{opt}) + \lim_{k \to \infty} [f(x_k)]^t(x^{opt} - x_k) =$$

(from Lemma 9)

$$\lim_{k \to \infty} [f(x_k)]^t[x - x^{opt}] + 0 =$$

(using Lemma 7)

$$[f(x^{opt})]^t(x - x^{opt}) \geq 0$$

due to the fact that x^{opt} is an optimal solution.

Putting everything back together we conclude that the limit exists, furthermore

$$\lim_{k \to \infty} f(x_k)^t(x - x_k) = [f(x^{opt})]^t(x - x^{opt}) \geq 0$$

for all $x \in K$. Therefore, every accumulation point x^* of the projection algorithm sequence (there exists at least one, since the limit $\lim_{k \to \infty} ||x_k - x^{opt}||$ exists and is finite) is indeed a VIP solution. In other words,

$$x^* \in K: \quad \lim_{k \to \infty} f(x_k)^t(x - x_k) = f(x^*)^t(x - x^*) \geq 0 \quad \forall x \in K.$$

Q.E.D.

We conclude this section by comparing the results we have obtained with the literature. As we have already observed, the projection algorithm fits into the framework of the general iterative scheme of Dafermos [12], the framework of linearization algorithms [32] and, when f is affine, the framework of the generalized contracting ellipsoid algorithm [18]. The choice of $g(x,y)$ in this case is $g(x,y) = \rho f(y) + G(x-y)$, with G positive definite and symmetric. Results in the literature have shown that the projection algorithm converges, as we've already stated in Section 2, under a norm condition, which in the special case of the projection algorithm becomes

$$||G^{-1/2}[G - \rho \nabla f(x)]G^{-1/2}|| < 1, \quad \forall x \in K.$$

On the other hand, when the problem function f is f-monotone, Theorem 3 establishes the convergence of the sequence of averages, while Theorem 4 establishes convergence of the sequence itself induced by the projection algorithm. In Section 2 we have shown that the f-monotonicity condition implies the previous norm condition on g, but in a less than or equal form.

Moreover, the convergence proof of the linearization algorithms and of the generalized contracting ellipsoid algorithm, under the norm condition, require an initial point close to the solution; in contrast, Theorems 3 and 4 establish global convergence. Therefore, under f-monotonicity, we obtain more general results.

4 Convergence of the relaxation scheme

In this section we show that when the problem function of a variational inequality (1) satisfies a norm condition, the sequence of averages induced by relaxation algorithms converges to an optimal solution x^{opt} of the problem (1). As in Section 3 (see Theorem 3), to establish this result, we employ the theory of variational inequalities as well as the ergodic theorem of Baillon. First, we need to describe the general relaxation scheme. It reduces the solution of the VIP (1) 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 $g : K \times K \rightarrow R^n$ satisfying the condition that

$$g(x,x) = f(x) \quad \forall x \in K.$$

We assume that for any fixed $y \in K$, the variational inequality

$$g(x', y)^t(x - x') \geq 0 \quad \forall x \in K \tag{6}$$

has a unique solution $x' \in K$, which can be computed by some known algorithm. As an important special case, $g(x, y)$ is defined by

$$g_i(x, y) = f_i(y^1, ..., y^{i-1}, x^i, y^{i+1}, ..., y^n) \quad i = 1, 2, ..., n.$$

Then solving (6) amounts to solving a separable variational inequality problem, which has a unique solution provided that g_i is a strictly monotone increasing function of the variable x^i. More generally, $g(x, y)$ should be defined so that the matrix $g_x(x, y)$ is symmetric and positive definite. In that case, as mentioned in Section 3, for a fixed value of $y \in K$ the variational inequality (6) is equivalent to a strictly convex minimization problem with the objective function $F(x) = \int g(x, y)dx$.

The Relaxation scheme

STEP 0:

Choose an arbitrary point $x_0 \in K$.

STEP $k+1$:

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

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

The original relaxation algorithms developed by Ahn and Hogan [1] 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 $i = 1, 2, ..., n$) to compute equilibria in economic equilibrium problems. This algorithm 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 [11] as well as the general variational inequality problem [12].

The papers [1], [11] and [12], and the references they cite in describe more details. In the subsequent analysis, we show that under appropriate assumptions, the limit of the sequence of averages induced by the sequence $\{x_k\}_{k=1}^{\infty}$ solves the original asymmetric variational inequality (1).

THEOREM 5:

Let $T : K \rightarrow R^n$ be the map which carries the point $y^* \in K$ to the solution $T(y^*) = x^*$ of the relaxation scheme (6) with $y = y^*$.

Suppose the algorithm function g satisfies the following conditions:

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

2. 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 the sequence of averages

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

converges to a solution of the VIP.

Proof:

We first prove that the map T, defined above, is a nonexpansive map in K. 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

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

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

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

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

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

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

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

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

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

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

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

Finally, the second assumption of this theorem, 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

$$\alpha||y_1 - y_2|| \geq \alpha||x_1 - x_2||. \quad (16)$$

Therefore,

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

Finally, the finite dimensional version of Theorem 2 (see Section 3) guarantees that the sequence of averages

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

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

Remarks:

1. The norm condition of Theorem 5, namely,

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

   implies the norm condition:

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

   analyzed in Section 2.

   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 inequality and

$$\alpha = \inf_{x,y \in K} (min\ eigenvalue\ g_x(x,y)) > 0.$$

2. The convergence proofs of Dafermos [11], [12] and of Ahn and Hogan [1] have more restrictive assumptions than the proof of Theorem 5. They require that the algorithm function g satisfies the following conditions:

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

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

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

We next give examples that violate these conditions but satisfy those of Theorem 5.

Examples:

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

$$M = \begin{bmatrix} 2 & 2 \\ -2 & 3 \end{bmatrix}.$$

This matrix is asymmetric, but positive definite since

$$M + M^t = \begin{bmatrix} 4 & 0 \\ 0 & 6 \end{bmatrix}$$

is positive definite. The matrix

$$g_x(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

is positive definite and symmetric. Moreover, $\alpha = \inf_{x,y \in K}$ (*min eigenvalue* $g_x(x, y)) = 2 > 0$, while

$$g_y(x, y) = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}.$$

Since $||g_y(x, y)||^2 = || \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix} ||^2 = \sup_{x \neq 0} \frac{x^t \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} x}{x^t x} = 4 = \alpha^2$.

Therefore, this problem satisfies the condition $||g_y(x, y)|| \leq \alpha$ of Theorem 5. To apply the norm condition of Dafermos and Ahn, Hogan, we would require $||g_y(x, y)|| \leq \lambda\alpha$ for some $0 < \lambda < 1$.

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

$$M = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}.$$

This matrix is symmetric and positive semidefinite. On the other hand, the matrix

$$g_x(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

is positive definite and symmetric. Moreover, $\alpha = \inf_{x,y \in K} (min\ eigenvalue\ g_x(x, y)) = 2 > 0$, while

$$g_y(x, y) = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}.$$

Since

$$\|g_y(x, y)\|^2 = \| \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix} \|^2 =$$

$$= \sup_{x \neq 0} \frac{x^t \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} x}{x^t x} = 4 = \alpha^2.$$

As in the case of Example 1, 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 $0 < \lambda < 1$.

5 On the convergence of the Frank-Wolfe algorithm

In this section we present a convergence proof of the Frank-Wolfe algorithm when applied to convex programming problems defined on general convex sets (i.e., variational inequality problems with a symmetric Jacobian matrix). This convergence proof applies to problems that satisfy the f-monotonicity condition. The convergence proof of Martos [28] assumes that the objective function F of the corresponding minimization problem is pseudoconvex, and that the feasible set is a bounded polyhedral. His main theorem establishes that every accumulation point of the sequence x_k induced by the algorithm is a solution to the minimization problem (or the equivalent variational inequality problem). Our convergence proof in this section also assumes symmetry of the Jacobian matrix, but permits the feasible set to be a general convex, compact set instead of a bounded polyhedron.

Before presenting the convergence proof, we make the following assumptions.

ASSUMPTIONS

A1. The feasible set K is a nonempty, convex and compact subset of R^n.

A2. The problem function f satisfies the f-monotonicity condition.

Since the Jacobian matrix ∇f is symmetric, the variational inequality problem is equivalent to the minimization problem:

$$min_{x \in K} \; F(x),$$

and, in this case, $f(x) = \nabla(F(x))$.

The Frank-Wolfe algorithm works as follows:

THE FRANK-WOLFE ALGORITHM

Step 0:

Choose an arbitrary point x_0 in the feasible set K. Set $k = 0$.

Step $k+1$:

Part a: (minimization part)

Let x_k be the point found in the previous step k and let y_{k+1} be the point that solves the following minimization problem:

$$\min_{y \in K} \; f(x_k)^t y. \tag{17}$$

(When the feasible set K is a polyhedron, this is a linear programming problem.)

TERMINATE if $f(x_k)^t(x_k - y_{k+1}) = 0$ and set x_k as the solution.

Part b: (line search part)

Find $x_{k+1} \in [x_k; y_{k+1}]$ for which

$$f(x_{k+1})^t(x - x_{k+1}) \geq 0 \quad \forall x \in [x_k; y_{k+1}] \tag{18}$$

Go to the next step $k + 2$, part a, with x_{k+1} in place of x_k.

The point x_{k+1} found at step $k + 1$ is a point that solves a one-dimensional variational problem. This is true because every point $x \in [x_k; y_{k+1}]$ can be rewritten as $x = x_k(a) = x_k + a(y_{k+1} - x_k)$ with $a \in [0, 1]$, while $x_{k+1} = x_k + a_{k+1}(y_{k+1} - x_k)$ with $a_{k+1} \in [0, 1]$. Therefore, we can write part b of the Frank Wolfe algorithm as: Find $a_{k+1} \in [0, 1]$ for which

$$f(x_k + a_{k+1}(y_{k+1} - x_k))^t(y_{k+1} - x_k)(a - a_{k+1}) \geq 0 \quad \forall a \in [0, 1]$$

which is the one-dimensional variational inequality $VI(h_k, [0, 1])$ with a problem function $h_k(a) = f(x_k(a))^t(y_{k+1} - x_k)$ and with the interval $[0, 1]$ as the feasible set. If the Jacobian matrix of the problem function f is symmetric and positive definite, with $f = gradient(F)$, part b is also equivalent to a one-dimensional minimization problem, namely

$$min_{a \in [0,1]} \; F(x_k(a)).$$

Then the VIP is also equivalent to the minimization problem

$$min_{x \in P} \; F(x).$$

To understand the behavior of $h_k(a)$, and therefore part b of step $k + 1$, we consider the following lemma.

LEMMA 10:

If the problem function f is strictly monotone, then the function $h_k : [0, 1] \rightarrow R$ defined as $h_k(a) = f(x_k(a))^t(y_{k+1} - x_k)$ is strictly increasing.

Proof:

To establish this result, we show that if $a_1 \neq a_2$, then

$$[h_k(a_1) - h_k(a_2)][a_1 - a_2] > 0.$$

This condition is true because

$$[h_k(a_1) - h_k(a_2)][a_1 - a_2] = [f(x_k(a_1)) - f(x_k(a_2))]^t(y_{k+1} - x_k)[a_1 - a_2] =$$

(recalling that $x = x_k(a) = x_k + a(y_{k+1} - x_k)$)

$$= [f(x_k(a_1)) - f(x_k(a_2))]^t [x_k(a_1) - x_k(a_2)] > 0,$$

when $a_1 \neq a_2$ (and therefore $x_k(a_1) \neq x_k(a_2)$), by the strict monotonicity of f. Therefore,

$$[h_k(a_1) - h_k(a_2)][a_1 - a_2] > 0.$$

Q.E.D.

Note that since $h_k(a)$ is strictly monotone, the one-dimensional variational inequality, find $a_{k+1} \in [0,1]$ so that $h_k(a_{k+1})^t(a - a_{k+1}) \geq 0$ for all $a \in [0,1]$, or equivalently (18), has a unique solution a_{k+1}, or $x_{k+1} = x_k + a_{k+1}(y_{k+1} - x_k)$, that satisfies one of two conditions.

1. If $h_k(1) = f(y_{k+1})^t(y_{k+1} - x_k) \leq 0$, then $a_{k+1} = 1$ since then $x_{k+1} = y_{k+1}$ and, therefore,

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

2. If $h_k(1) > 0$, then $a_{k+1} \in [0,1]$ and $h_k(a_{k+1}) = 0$.
   This result is true because $h_k(0) = f(x_k)^t(y_{k+1} - x_k) < 0$
   (from part a, $y_{k+1} = argmin_{y \in P_k} \ f(x_k)^t y$, also $h_k(0) \neq 0$ otherwise we would stop),
   and h_k is a strictly increasing function (Lemma 10), $h_k(1) > h_k(0) < 0$. Therefore, either $h_k(1) \leq 0$ so $a_{k+1} = 1$, or $h_k(1) > 0$ and so $h_k(a_{k+1}) = 0$, for some $a_{k+1} \in [0,1]$.

If f is strictly monotone, we can view part b (because of Lemma 10) as a minimization problem, regardless of the symmetry of the Jacobian matrix of f. Namely,

$$min_{a \in [0,1]} \ (-h_k(a))^2. \tag{19}$$

Remark:

When $h_k(1) > 0$, computing a_{k+1} boils down to solving the one variable equation $h_k(a_{k+1}) = 0$, which becomes

$$h_k(a_{k+1}) = f(x_k + a_{k+1}(y_{k+1} - x_k))^t(y_{k+1} - x_k) = 0$$

$$\text{or}\quad [f(x_k + a_{k+1}(y_{k+1} - x_k)) - f(x^k)]^t(y_{k+1} - x_k) = f(x_k)^t(x_k - y_{k+1}).$$

Applying the mean value theorem shows that for some $z \in [x_k; x_{k+1}]$,

$$f(x_k)^t(x_k - y_{k+1}) = a_{k+1}||y_{k+1} - x_k||^2_{\nabla f(z)}, \text{ implying that}$$

$$x_{k+1} = x_k + \frac{f(x_k)^t(x_k - y_{k+1})}{||x_k - y_{k+1}||^2_{\nabla f(z)}}(y_{k+1} - x_k),$$

for some $z \in [x_k; x_{k+1}]$.

We are now ready to establish convergence of the algorithm sequence x_k. Before we state the main theorem, we establish several preliminary lemmas.

LEMMA 11:

Let x_k be the sequence induced by the Frank-Wolfe algorithm. Under assumptions A1 and A2, the objective function $\{F(x_k)\}$ is a convergent sequence.

Proof:

Observe that under f-monotonicity,

$$F(x_k) - F(x_{k+1}) \geq \nabla F(x_{k+1})^t(x_k - x_{k+1}) \geq 0,$$

(this inequality is due to the line search we performed in part b of the algorithm). Therefore, $F(x_k) \geq F(x_{k+1})$ is a decreasing sequence. Moreover, since the feasible set is bounded, the sequence $F(x_k)$ is bounded from below. So $\{F(x_k)\}_{k=0}^{\infty}$ is a convergent sequence. Q.E.D.

LEMMA 12:

Let x_k be the sequence induced by the Frank-Wolfe algorithm. Under assumptions A1 and A2, every convergent subsequence x_{k_p} (with accumulation point x^*) has (perhaps a further) convergent subsequence y_{k_p+1} (with accumulation point y^*).

Proof:

This follows from assumption A1, i.e., that the feasible set K is bounded. Q.E.D.

LEMMA 13:

Let x_k be the sequence induced by the Frank-Wolfe algorithm. Under assumptions A1 and A2,

$$\lim_{k_p \to \infty} f(x_{k_p})^t (x_{k_p} - y_{k_p+1}) = 0$$

Proof:

Consider any convergent subsequence x_{k_p} with accumulation point x^*. From Lemma 12, the subsequence y_{k_p+1} has an accumulation point y^*.

Let $s, r \in \{k_p\}_{k=0}^{\infty}$ with $s \geq r+1 > r$. Then $F(x_s) \leq F(x_{r+1}) \leq F(x_r)$ and

$$F(x_s) \leq F(x_{r+1}) \leq F(x_r + a(y_{r+1} - x_r)), \quad \forall a \in [0,1].$$

Therefore $\frac{F(x_s)-F(x_r)}{a} \leq \frac{F(x_r + a(y_{r+1}-x_r)) - F(x_r)}{a}$ $\forall a \in [0,1]$. Letting $s \to \infty$ and $r \to \infty$ in both sides of the inequality, we see that

$$\frac{F(x^* + a(y^* - x^*)) - F(x^*)}{a} \geq 0 \ \ \forall a \in [0,1].$$

Letting $a \to 0$, we see that $f(x^*)^t(y^* - x^*) \geq 0$. Since $f(x_{k_p})^t(x_{k_p} - y_{k_p+1}) \geq 0$ from part a of step $k_p + 1$, we also conclude that $f(x^*)^t(y^* - x^*) \leq 0$. Finally, combining this result with the previous inequality, we conclude that $f(x^*)^t(y^* - x^*) = 0$. Q.E.D.

LEMMA 14:

Let x_k be the sequence induced by the Frank-Wolfe algorithm and let x^* be any accumulation point of sequence x_k. Under assumptions A1 and A2,

$$f(x^*) = f(x^{opt}) \ \text{ and } \ f(x^*)^t(x^{opt} - x^*) = 0.$$

Proof:

Part a of step $k_p + 1$ and the fact that $x^{opt} \in K$ is an optimal solution implies that

$$f(x_{k_p})^t(y_{k_p+1} - x_{k_p}) \leq f(x_{k_p})^t(x^{opt} - x_{k_p}) \leq 0.$$

Lemma 13 implies that $f(x_{k_p})^t(x^{opt} - x_{k_p}) \longrightarrow_{k_p \to \infty} 0$, i.e.,

$$f(x^*)^t(x^{opt} - x^*) = 0$$

for any accumulation point x^* of the sequence x_k. Combined with f-monotonicity, the result implies that

$$0 = f(x^*)^t(x^* - x^{opt}) \geq (f(x^*) - f(x^{opt})^t(x^* - x^{opt}) \geq a||f(x^*) - f(x^{opt})||^2.$$

Therefore, $f(x^*) = f(x^{opt})$. Q.E.D.

THEOREM 6:

Let x_k be the sequence induced by the Frank-Wolfe algorithm. Under assumptions A1 and A2, every accumulation point x^* of the sequence x_k is a VIP solution.

Proof:

In Lemma 14 we have shown that

$$f(x^*) = f(x^{opt}) \text{ and } f(x^*)^t(x^{opt} - x^*) = 0.$$

These two equalities imply that

$$f(x^*)^t(x-x^*) = f(x^*)^t(x-x^{opt})+f(x^*)^t(x^{opt}-x^*) = f(x^*)^t(x-x^{opt}) = f(x^{opt})^t(x-x^{opt}) \geq 0,$$

for all $x \in K$. Therefore, x^* is a VIP solution.

Q.E.D.

6 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,

$$d_w = \sum_{p \text{ joining } w} F_p, \quad \forall\, O-D \;\; pair\ w, \tag{20}$$

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 by

$$f_i = \sum_{p \text{ passing through } i} F_p, \tag{21}$$

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,

$$F_p \geq 0 \;\; \forall \text{ paths } \; p, \tag{22}$$

induces the link flow f through (21) and is connected to the demand vector d through (20). 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:

$$C_p(F) = \sum_{i \in path\ p} c_i(f), \;\; \forall \text{ paths } \; p. \tag{23}$$

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

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

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

$$f^* \in K \text{ is user optimized if and only if } c(f^*)^t(f - f^*) \geq 0, \quad \forall f \in K. \qquad (24)$$

Several papers [10], [11], [25], [36], [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 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 Section 2, 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 3 and 5 guarantee that the sequence of averages induced by the projection and the relaxation 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, Theorem 4 establishes that every accumulation point of the sequence induced by the projection algorithm is a user optimizing load pattern.

Next, we study some traffic equilibrium examples that illustrate the importance of the f-monotonicity condition.

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

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

This matrix is positive semidefinite if $g_i - a g_i^2 \geq 0$ for $i = 1, 2, ..., n$.

This, in turn, is true if the congestion coefficients $g_i \geq 0$ for $i = 1, 2, ..., n$ and $a \leq \frac{1}{max_{1 \leq i \leq n} g_i}$.

The matrix is positive definite, and so the function c is strongly monotone, if each $g_i > 0$ and $a < \frac{1}{max_{1 \leq i \leq n} g_i}$. It is positive semidefinite, and so (from Section 2) is f-monotone even if some $g_i = 0$. Our analysis still applies even though some or all $g_i' s$ are zero. This example shows that f-monotonicity might permit some links of the network to be uncongested. This might very well be the case in large scale networks. The projection algorithm would still allow us, as shown in Theorem 3 and Theorem 4, to compute an optimal solution to the problem.

2. Consider the simple case of a transportation network, with one O-D pair $w = (x, y)$ and three links connecting this O-D pair as shown in Figure 1. In this case, the link congestion function is not separable.

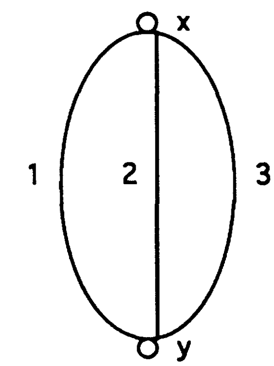

Figure 1: The traffic equilibrium problem

The travel costs on the links are

$$c_1(f) = f_1 + f_2 + 10,$$

$$c_2(f) = \frac{1}{2}f_1 + 2f_2 + 5,$$

$$c_3(f) = 15.$$

Suppose the demand for the O-D pair w is $d_w = 20$. The user equilibrium solution is $f_1 = 0,\ f_2 = 5 > 0,\ f_3 = 15 > 0$. At this point, $c_1 = c_2 = c_3$. In this case,

$$\nabla c = M = \begin{bmatrix} 1 & 1 & 0 \\ \frac{1}{2} & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$M^t = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$M + M^t = \begin{bmatrix} 2 & \frac{3}{2} & 0 \\ \frac{3}{2} & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which is positive semidefinite, but not positive definite. Furthermore,

$$M^t - aM^tM = \begin{bmatrix} 1-\frac{5}{4}a & \frac{1}{2}-2a & 0 \\ 1-2a & 2-5a & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

for $a = 1/5$, becomes

$$M^t - \frac{1}{5}M^tM = \begin{bmatrix} \frac{3}{4} & \frac{1}{10} & 0 \\ \frac{3}{5} & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which is a positive semidefinite matrix, since its symmetric part is

$$\begin{bmatrix} \frac{3}{2} & \frac{7}{10} & 0 \\ \frac{7}{10} & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(which is indeed positive semidefinite). In this example, the travel cost function c is f-monotone, but not strongly monotone.

3. We conclude this set of examples by considering a transportation network with multiple equilibria, specifically 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 $f_1 + f_2 = 20, \quad f_3 = 0$ is a solution to the user optimized problem, since then $c_1 = c_2 = 25 \leq c_3 = 30$.

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 f-monotone, but not strongly monotone.

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

LEMMA 15:

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

$$[\sum_{i=1}^{n}(x_i)]^2 \leq n\sum_{i=1}^{n}(x_i)^2. \tag{25}$$

This result is easy to establish by induction.

Proposition 2:

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 f-monotone with respect to the constant a, then the path cost function $C = C(F)$ is also f-monotone with respect to the constant $a' = \frac{a}{nN}$.

Proof:

If the link cost function $c = c(f)$ is 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 $C_p(F) = \sum_{i \in \text{path } p} c_i(f)$. 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 (23) and Lemma 15 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 f-monotone. Q.E.D.

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

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

Therefore, 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.

As our prior results show, we can consider networks that contain some uncongested paths. A path flow algorithm is preferable to a link flow one because it is much less costly to carry out projection iterations in the space of path flows than in the space of link flows [6].

Proposition 2 is similar to a result of Bertsekas and Gafni, [6]: they assume strong (rather than f-) monotonicity on the link cost function and show that a path flow projection algorithm solves the user optimizing equilibrium problem.

7 Conclusions and open questions

In this paper, we analyzed the convergence properties of several classical algorithms — the Frank-Wolfe algorithm and projection and relaxation algorithms — with respect to the condition of f-monotonicity which is weaker than the standard strong monotonicity condition. We began by showing the connection between f-monotonicity and the norm condition of Dafermos [12], Chan and Pang [32], and Hammond and Magnanti [18]. Assuming the f-monotonicity condition, we showed that the sequence of averages induced by the projection algorithm converges to a solution of the variational inequality problem. Under a norm condition weaker than an existing one, we also established the convergence of the sequence of averages induced by relaxation algorithms. To establish these two results, we employed an ergodic theorem for nonexpansive maps due to Baillon [3]. Moreover, we showed that when the feasible set K is both compact and convex, every accumulation point of the projection algorithm solves the variational inequality problem. We also showed that every accumulation point of the sequence induced by the Frank-Wolfe algorithm, when applied to convex optimization problems over convex sets, is a VIP solution under the f-monotonicity condition. Finally, we applied these results to transportation networks, permitting uncongested links. We showed that a path flow projection algorithm can be used to solve the user optimizing problem when the link cost function is f-monotone.

The results in this paper suggest the following question:
can some form of the 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] C. Baiocchi and A. Capelo (1984), *Variational and Quasivariational Inequalities: Application to Free Boundary Problems*, Wiley, New York.

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

[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] M. Bruynooghe, A. Gilbert and M. Sakarovitch (1968), "Une methode d' affectation du traffic", *in Fourth Symposium of the theory of traffic flow*, Karlsruhe.

[9] E.W. Cheney and A.A. Goldstein (1959), "Newton's method for convex programming and Tchebycheff approximation", *Numerische Math.*, 1, 253-268.

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

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

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

[13] M. Frank and P. Wolfe (1956), "An algorithm for quadratic programming", *Naval Research Logistic Quarterly*, 3, 95-110.

[14] B. Golden (1975), "A minimum cost multi-commodity network flow problem concerning imports and exports", *Networks* 5, 331-356.

[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] J.H. Hammond and T.L. Magnanti (1987), "A contracting ellipsoid method for variational inequality problems", Working paper OR, 160-87, Operations Research Center, M.I.T.

[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] L.J. Leblanc, E.K. Morlok and W.P. Pierskalla (1975), "An efficient approach to solving the road network equilibrium traffic assignment problem", *Transportation Research*, 5, 309-318.

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

[23] J.L. Lions and G. Stampacchia (1967), "Variational inequalities", *Communications on Pure and Applied Mathematics*, 20, 493-519.

[24] D.G. Luenberger (1973), *Linear and nonlinear programming*, Addison Wesley, Reading, M.A.

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

[26] 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.

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

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

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

[30] S. Nguyen and C. Dupuis (1984), "An efficient method for computing traffic equilibria in networks with asymmetric transportation costs", *Transportation Science*, 18, 185-202.

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

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

[33] 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.

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

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

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

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

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

[39] W.I. Zangwill (1969), *Nonlinear Programming: A Unified Approach*, Prentice-Hall, Englewood Cliffs, N.J.

[40] S.I. Zuhovickki, R.A. Polyak, and M.E. Primak (1969), "Two methods of search for equilibrium points of n-person concave games", *Soviet Mathematics Doklady*, 10:2, 279-282.