OPERATIONS RESEARCH CENTER

Working Paper

Solving Variational Inequality and Fixed Point Problems by Averaging and Optimizing Potentials

by

T.L. Magnanti
G. Perakis

OR 324-97 December 1997

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Solving Variational Inequality and Fixed Point Problems by Averaging and Optimizing Potentials

by

T.L. Magnanti

G. Perakis

OR 324-97 December 1997

Solving Variational Inequality and Fixed Point Problems by Averaging and Optimizing Potentials

Thomas L. Magnanti * and Georgia Perakis †

December 1997 ‡

Abstract

We introduce a general adaptive averaging framework for solving fixed point and variational inequality problems. Our goals are to develop schemes that (i) compute solutions when the underlying map satisfies properties weaker than contractiveness, (for example, weaker forms of nonexpansiveness), (ii) are more efficient than the classical methods even when the underlying map is contractive, and (iii) unify and extend several convergence results from the fixed point and variational inequality literatures. To achieve these goals, we consider line searches that optimize certain potential functions. As a special case, we introduce a modified steepest descent method for solving systems of equations that does not require a previous condition from the literature (the square of the Jacobian matrix is positive definite). Since the line searches we propose might be difficult to perform exactly, we also consider inexact line searches.

Key Words:

Fixed Point Problems, Variational Inequalities, Averaging Schemes, Nonexpansive Maps, Strongly-f-Monotone Maps.

---

*Sloan School of Management and Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139.

†Operations Research Center, Cambridge, MA 02139.

‡Preparation of this paper was supported, in part, by NSF Grant 9634736-DMI from the National Science Foundation.

1 Introduction

Fixed point and variational inequality problems define a wide class of problems that arise in optimization as well as areas as diverse as economics, game theory, transportation science, and regional science. Their widespread applicability motivates the need for developing and studying efficient algorithms for solving them.

Often algorithms for solving problems in these various settings establish an algorithmic map $T : K \subseteq R^n \rightarrow K$ is a given map defined over a closed, convex (constraint) set K in R^n whose fixed point solution solves the original problem.

$$FP(T,K): \quad \text{Find } \; x^* \in K \subseteq R^n \text{ satisfying } \; T(x^*) = x^*. \tag{1}$$

The algorithmic map T might be defined through the solution of a subproblem. For variational inequality problems, examples of an algorithmic map T include a projection operator (see Proposition 1) or the solution of a simpler variational inequality subproblem (see, for example, [12] and [44]).

A standard method for solving fixed point problems for contractive maps $T(.)$ in R^n is to apply the iterative procedure

$$x_{k+1} = T(x_k). \tag{2}$$

The classical Banach fixed point theorem shows that this method converges from any starting point to the unique fixed point of T. When the map is nonexpansive instead of contractive, this algorithm need not converge and, indeed, the map T need not have a fixed point (or it might have several). For example, the sequence that the classical iteration (2) induces for the 90^o degree rotation mapping shown in Figure 1 does not converge to the solution x^*.

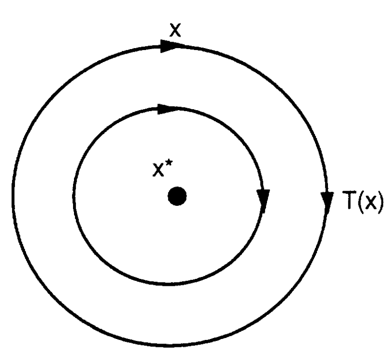

Figure 1: 90^o Degree Rotation

In these situations, researchers (see [2], [10], [13], [27], [38], [39], [46]) have established convergence of recursive averaging schemes of the type

$$x_{k+1} = x_k + a_k(T(x_k) - x_k),$$

assuming that $0 < a_k < 1$ and $\sum_{k=1}^{\infty} a_k(1 - a_k) = +\infty$. Equivalently, this condition states that $\sum_{k=1}^{\infty} \min(a_k, 1 - a_k) = +\infty$, implying that the iterates lie far "enough" from the previous iterates as well as from the image of the previous iterates under the fixed point mapping. For variational inequality problems, averaging schemes of this type give rise to convergent algorithms using algorithmic maps T that are nonexpansive rather than contractive (see Magnanti and Perakis [38]).

Even though recursive averaging methods converge to a solution whenever the underlying fixed point map is nonexpansive, they might converge very slowly. In fact, when the underlying map T is a contraction, recursive averaging might converge more slowly than the classical iterative method (2).

Motivated by these observations, in this paper we introduce and study a recursive averaging framework for solving fixed point and variational inequality problems. Our goals are to (i) design methods permitting a larger range of step sizes with better rates of convergence than the classical iterative method (2) even when applied to contractive maps, (ii) impose assumptions on the map T that are weaker than contractiveness, (iii) understand the role of nonexpansiveness, and (iv) unify and extend several convergence results from the fixed point and variational inequality literature. To achieve these goals, we introduce a general adaptive averaging framework that determines step sizes by "intelligently" updating them dynamically as we apply the algorithm. For example, to determine step sizes for nonexpansive maps, we adopt the following idea: unless the current iterate lies close to a solution, the next iterate should not lie close to either the current iterate or its image under the fixed point map T. As a special case, we introduce a modification of the classical steepest descent method for solving fixed point problems (or equivalently, unconstrained asymmetric variational inequality problems). This new method does not require the square of the Jacobian matrix to be positive definite as does the classical steepest descent method, but rather that the Jacobian matrix be positive semidefinite. Moreover, it is a globally convergent method even in the general nonlinear case.

1.1 Preliminaries: Fixed Points and Variational Inequalities

Fixed point problems are closely related to variational inequality problems

$$VI(f,K): \quad \text{Find } x^* \in K \subseteq R^n : \ f(x^*)^t(x - x^*) \geq 0, \ \forall x \in K \tag{3}$$

defined over a closed, convex (constraint) set K in R^n. In this formulation, $f : K \subseteq R^n \to R^n$ is a given function.

The following well-known proposition provides a connection between the two problems. In stating this result and in the remainder of this paper, we let G be a positive definite and symmetric matrix. We also let Pr_K^G denote the projection operator on the set K with respect to the norm $\|x\|_G = (x^t G x)^{\frac{1}{2}}$.

Proposition 1 *([25]): Let ρ be a positive constant and T be the map $T = Pr_K^G(I - \rho G^{-1} f)$. Then the solutions of the fixed point problem $FP(T,K)$ are the same as the solutions of the variational inequality problem $VI(f,K)$, if any.*

Corollary 1 *: Let $f = I - T$. Then the fixed point problem $FP(T,K)$ has the same solutions, if any, as the variational inequality problem $VI(f,K)$.*

To state the results in this paper, we need to impose several conditions on the underlying fixed point map T or the variational inequality map f. We first introduce a definition that captures many well known concepts from the literature.

Definition 1 *: A map $T : K \to K$ is A-**domain** and B-**range coercive** on the set K with coercive constants $A \in R$ and $B \in R$ if*

$$\|T(x) - T(y)\|_G^2 \leq A\|x - y\|_G^2 + B\|x - T(x) - y + T(y)\|_G^2 \quad \forall x, y \in K. \tag{4}$$

T is a **(range) coercive** *map on K relative to the $\|.\|_G$ norm if $A > 0$ and $B = 0$.*
T is a **(range) contraction on** *K, relative to the $\|.\|_G$ norm, if $0 < A < 1$ and $B = 0$*
and is **(range) nonexpansive on** *K, relative to the $\|.\|_G$ norm, if $A = 1$ and $B = 0$.*
T is **nonexpansive-contractive** *if $A = 1$ and $B < 1$,*
T is **contractive-nonexpansive** *if $A < 1$ and $B = 1$,*
T is **nonexpansive-nonexpansive** *if $A = 1$ and $B = 1$,*
T is **firmly contractive** *on K relative to the $\|.\|_G$ norm if $0 < A < 1$ and $B = -1$,*
and T is **firmly nonexpansive** *on K relative to the $\|.\|_G$ norm if $A = 1$ and $B = -1$.*

Observe that any firmly contractive map is a contraction and any firmly nonexpansive map is nonexpansive.

When y in the definition of a nonexpansive map is restricted to be a fixed point solution, researchers have referred to the map T as *pseudo nonexpansive* relative to the $\|.\|_G$ norm (see [55]).

Observe that any nonexpansive map is also nonexpansive-contractive. The following example shows that the converse isn't necessarily true.

Example:

If $T(x) = -3x$ then $x^* = 0$. T is an expansive map about x^* since $\|T(x) - T(x^*)\| > \|x - x^*\|$ for all x. Nevertheless, the map T satisfies the *nonexpansive-contractive* condition, with $B = \frac{1}{2}$, since

$$9\|x\|^2 = \|T(x) - T(x^*)\|^2 = \|x - x^*\|^2 + B\|x - T(x)\|^2 = 9\|x\|^2.$$

Definition 2 : *For variational inequality problems $VI(f, K)$, the following notions are useful.*
*f is b-**domain and** d-**range monotone** (on the set K) with monotonicity constants b and d if*

$$[f(x) - f(y)]^t G[x - y] \geq b\|x - y\|_G^2 + d\|f(x) - f(y)\|_G^2, \quad \forall x, y \in K.$$

As special cases:

(i) f is *strongly domain and range monotone* if $b > 0$ and $d > 0$.

(ii) f is *strongly monotone* if $b > 0$ and $d = 0$.

(iii) f is *strongly-f-monotone* if $b = 0$ and $d > 0$.

(iv) f is *monotone* if $b = d = 0$.

Contraction, nonexpansiveness, firm nonexpansiveness, pseudo nonexpansiveness, monotonicity, and strong monotonicity are standard conditions in the literature (especially when $G = I$). Some authors refer to strong monotonicity and strong-f-monotonicity as coercivity and co-coercivity (see [54], [57]).

Lemma 1 : *Let $T : R^n \to R^n$ and $f : R^n \to R^n$ be two given maps with $T = I - f$. Then T is domain and range coercive with coercivity constants A and B if and only if f is domain and range monotone with monotonicity constants $b = \frac{1-A}{2}$ and $d = \frac{1-B}{2}$. (We permit the coercivity and monotonicity constants to be negative.) That is,*

$$\|T(x) - T(y)\|_G^2 \leq A\|x - y\|_G^2 + B\|x - T(x) - y + T(y)\|_G^2$$

if and only if

$$[f(x) - f(y)]^t G[x - y] \geq \frac{1-A}{2}\|x - y\|_G^2 + \frac{1-B}{2}\|f(x) - f(y)\|_G^2.$$

Proof: Substitute $I - f = T$, expand $\|T(x) - T(y)\|_G^2$, and rearrange terms.

Corollary 2 *:*

1. *T is nonexpansive relative to the $\|.\|_G$ norm if and only if f is strongly-f-monotone relative to the matrix G with a monotonicity constant $d = \frac{1}{2}$.*

2. *If T is contractive relative to the $\|.\|_G$ norm, then f is strongly monotone relative to the matrix G with a monotonicity constant $b = \frac{1-A}{2}$.*

3. *If T is nonexpansive-nonexpansive (nonexpansive-contractive) relative to the $\|.\|_G$ norm with a coercivity constant $B \in [0,1]$ ($B < 1$ in the nonexpansive-contractive case) if and only if f is strongly-f-monotone with constant $d = \frac{1-B}{2}$.*

4. *T is firmly contractive relative to the $\|.\|_G$ norm with constant $A \in [0,1)$ if and only if f is domain and range monotone with constants $b = \frac{1-A}{2}$ and $d = 1$.*

Proof of the Corollary: Take $A = 1$ and $B = 0$ for part a, $A < 1$ and $B = 0$ for part b, $A = 1$ and $B = 1$ (or $B < 1$) in part c, and $B = -1$ in part d.

The following related result is also useful.

Proposition 2 *([8]): Let $\rho > 0$ be a given constant, G a positive definite and symmetric matrix, and $f : K \subseteq R^n \to R^n$ a given function. Then the map $T = I - \rho G^{-1} f$ is nonexpansive with respect to the $\|.\|_G$ norm if and only if f is strongly-f-monotone.*

We structure the remainder of this paper as follows. In Section 2 we introduce a general adaptive framework for solving fixed point problems that evaluates step sizes through the optimization of a potential. In Section 3 we introduce and study several choices of potentials as special cases. At each step, we optimize these potentials to dynamically update the step sizes of the recursive averaging scheme. We also examine rates of convergence for the various choices of potentials. To make the computation of the step sizes easier to perform, in Section 4 we attempt to understand when the schemes have a better rate of convergence than the classic iteration (2). In Section 5 we consider inexact line searches. Finally, in Section 6 we address some open questions.

2 Potential Optimizing Methods: A General Case

2.1 An Adaptive Averaging Framework

We are interested in finding "good" step sizes a_k for the following general iterative scheme

$$x_{k+1} = x_k(a_k) = x_k + a_k(T(x_k) - x_k).$$

For any positive definite and symmetric matrix G, the fixed point problem $FP(T, K)$ is equivalent to the minimization problem

$$min_{x \in K} \|x - T(x)\|_G^2. \tag{5}$$

The difficulty in using the equivalent optimization formulation (5) is that even when T is a contractive map relative to the $\|.\|_G$ norm, the potential $\|x - T(x)\|_G^2$ need not be a convex function.

Example:

Let $T(x) = x^{1/2}$ and $K = [1/2, 1]$. Then the fixed point problem becomes

$$\text{Find } x^* \in [1/2, 1] : (x^*)^{1/2} = x^*.$$

$x^* = 1$. The mapping T is contractive on K since

$$\|T(x) - T(y)\| = \|x^{1/2} - y^{1/2}\| = \frac{\|x - y\|}{\|x^{1/2} + y^{1/2}\|} \leq \frac{\sqrt{2}}{2}\|x - y\|,$$

for all $x, y \geq 1/2$. The potential $g(x) = (x - T(x))^2 = (x - x^{1/2})^2$ is not convex for all $x \in [1/2, 9/16)$. In fact, since $g(x)'' = \frac{4 - 3\frac{1}{\sqrt{x}}}{2} < 0$ for all $x < 9/16$, $g(x)$ is concave in the interval $[1/2, 9/16)$.

This example shows that the minimization problem (5) need not be a convex programming problem. Nevertheless, the equivalent minimization problem does motivate a class of potential functions and a general scheme for computing the step sizes a_k.

Adaptive Averaging Framework

$$a_k = argmin_{\{a \in S\}} g(x_k(a)).$$

Later, we consider several choices for the potential $g(x_k(a))$. We assume the step size search set S is a subset of R. Examples include $S = [0, 1]$, $S = R^+$ and $S = [-1, 0]$.

We first make several observations, beginning with a relationship between $\|x - T(x)\|_G^2$ and $\|x - x^*\|_G^2$.

Preliminary Observations

Proposition 3 : *If T is a coercive map relative to the $\|.\|_G$ norm with a coercive constant $A > 0$ and if $x \neq x^*$, then the following inequality is valid,*

$$(1+\sqrt{A})^2 \geq \frac{\|x - T(x)\|_G^2}{\|x - x^*\|_G^2} \geq (1-\sqrt{A})^2. \tag{6}$$

Proof: Lemma 1 with $y = x^*$ and $B = 0$ together with the inequality $|(x - T(x))^t G(x - x^*)| \leq \|x - T(x)\|_G . \|x - x^*\|_G$ imply that

$$\|x - T(x)\|_G . \|x - x^*\|_G \geq \frac{1-A}{2}\|x - x^*\|_G^2 + \frac{1}{2}\|x - T(x)\|_G^2.$$

Dividing by $\|x - x^*\|_G^2$ on both sides and setting $y = \frac{\|x-T(x)\|_G}{\|x-x^*\|_G}$, we obtain $\frac{1}{2}y^2 - y + \frac{1-A}{2} \leq 0$. This inequality holds for the values of y lying between the two roots $y_1 = 1 + \sqrt{A}$ and $y_2 = 1 - \sqrt{A}$ of the binomial. This result implies the inequality (6). Q.E.D.

Definition 3 : *The sequence $\{x_k\}$ is asymptotically regular (with respect to the map T) if* $\lim_{k\to\infty} \|x_k - T(x_k)\| = 0$.

Proposition 4 :

i. If T is a continuous map and the sequence $\{x_k\}$ converges to some fixed point x^*, then the sequence $\{x_k\}$ is asymptotically regular.

ii. If the sequence $\{x_k\}$ is bounded and asymptotically regular, then every limit point of this sequence is also a fixed point solution.

Proof: Property (i) is a direct consequence of continuity since if the sequence $\{x_k\}$ converges to some fixed point x^*, then the continuity of T implies that $\lim_{k\to+\infty} \|x_k - T(x_k)\|_G^2 = 0$.

Result (ii) follows from the observation that if the sequence $\{x_k\}$ is bounded, then it has at least one limit point. Asymptotic regularity then implies that every limit point is also a fixed point solution. Q.E.D.

Remark: If the sequence $\|x_k - x^*\|$ is convergent for *every* fixed point solution x^*, then the entire sequence $\{x_k\}$ converges to a fixed point solution.

This remark follows from property (ii) of Proposition 4 since every limit point $\bar{x}$ is also a fixed point solution x^* and the sequence $\|x_k - x^*\|_G$ is convergent for every limit point $x^* = \bar{x}$. This result further implies that the limit point $\bar{x}$ of the sequence $\{x_k\}$ must be unique. Therefore, the entire sequence $\{x_k\}$ converges to a fixed point solution.

Corollary 3 *([38]): If T is a nonexpansive map and for each $k = 1, 2, \ldots$, $x_{k+1} = x_k + a_k(T(x_k) - x_k)$, with $a_k \in [0, 1]$, then the following two statements are equivalent:*

i. For some fixed point x^*, $\lim_{k \to +\infty} \|x_k - x^*\|_G^2 = 0$.

ii. The sequence $\{x_k\}$ is asymptotically regular.

Corollary 4 *([38]): Let T be any coercive map. Let $x_k(a) = x_k + a(T(x_k) - x_k)$ for any $0 \leq a \leq 1$ and let x^* be any fixed point of T. Then,*

i. $\|x_k(a) - x^*\|_G \leq [1 - a(1 - \sqrt{A})]\|x_k - x^*\|_G$, and

ii. $\|x_k(a) - T(x_k(a))\|_G \leq [1 - a(1 - \sqrt{A})]\|x_k - T(x_k)\|_G$.

Remark:

Corollary 4 implies that for any potential function $g(x_k(a))$ with $a \in [0, 1]$, if T is nonexpansive and x^* is any fixed point of T, then the iterates x_k of the general scheme satisfy the following conditions:

i. $\|x_k - x^*\|_G$ is nonincreasing and, therefore, a convergent sequence.

ii. $\|x_k - T(x_k)\|_G$ is nondecreasing and, therefore, a convergent sequence.

In analyzing the convergence of the adaptive averaging framework, we will impose several conditions on the map T and on the step sizes $\{a_k\}$ in the iterative scheme $x_{k+1} = x_k + a_k(T(x_k) - x_k)$. For this purpose, we will select the step sizes from a class C of sequences $\{a_k\}$. For example, the class C might be the set of all sequences generated by a family of potential functions g, that is, $a_k = argmin_{a \in S} g(x_k(a))$, for some step size search set S, some potential function g and $x_k(a) = x_k + a(T(x_k) - x_k)$. Alternatively, C might be the set of all *Dunn sequences*, that is, sequences satisfying the conditions $0 < a_k < 1$ and $\sum_{k=1}^{\infty} a_k(1 - a_k) = +\infty$.

Convergence Analysis

For some positive definite, symmetric matrix $\bar{G}$ and any sequence of step sizes $\{a_k\}$ from class C, for a fixed point solution x^* of the map T, we define the quantity

$$A_k(x^*) \equiv \frac{\|x_k - x^*\|_{\bar{G}}^2 - \|T(x_k) - T(x^*)\|_{\bar{G}}^2}{\|x_k - T(x_k)\|_{\bar{G}}^2} + (1 - a_k).$$

Observe that if T is nonexpansive, then the first term in the definition of $A_k(x^*)$ is nonnegative. If, additionally, $0 \leq a_k \leq 1$, then the quantity $A_k(x^*)$ is nonnegative. Therefore, when $0 \leq a_k \leq 1$, the condition $A_k(x^*) \geq 0$ provides a generalization of the condition T is a nonexpansive map.

We use a generalization of this condition by imposing the following assumptions in our convergence analysis.

A1. For any fixed point x^* of the map T, $a_k A_k(x^*)$ is nonnegative.

A2. For any fixed point x^* of the map T, if $a_k A_k(x^*)$ converges to zero, then every limit point of the sequence $\{x_k\}$ is a fixed point solution.

As we have already observed, whenever T is nonexpansive and $0 \leq a_k \leq 1$, assumption A1 is valid.

The following theorem establishes a convergence result for the general adaptive averaging framework.

Theorem 1 *: Consider iterates of the type $x_{k+1} = x_k + a_k(T(x_k) - x_k)$, with step sizes a_k chosen from a class C. Assume that the map T has a fixed point solution x^*. If the map T and the sequence $\{a_k\}$ in the class C generated by the adaptive averaging framework satisfy conditions A1 and A2, then the sequence of iterates $\{x_k\}$ converges to a fixed point solution.*

Proof: Let $F(x) = x - T(x)$. Consider any fixed point x^* of map T. Observe that

$$A_k(x^*) = \frac{\|x_k - x^*\|_{\bar{G}}^2 - \|T(x_k) - T(x^*)\|_{\bar{G}}^2}{\|x_k - T(x_k)\|_{\bar{G}}^2} + (1 - a_k) = \frac{2(F(x_k) - F(x^*))^t \bar{G}(x_k - x^*)}{\|F(x_k) - F(x^*)\|_{\bar{G}}^2} - a_k.$$

Suppose $F(x_k) \neq 0$. Since x^* is a fixed point solution of the map T, $F(x^*) = 0$ and, therefore,

$$\|x_{k+1} - x^*\|_{\bar{G}}^2 = \|x_k - a_k F(x_k) - x^*\|_{\bar{G}}^2 =$$

$$\|x_k - x^*\|_{\bar{G}}^2 + a_k^2 \|F(x_k) - F(x^*)\|_{\bar{G}}^2 - 2a_k(F(x_k) - F(x^*))^t \bar{G}(x_k - x^*).$$

$$\|x_{k+1} - x^*\|_{\bar{G}}^2 = \|x_k - x^*\|_{\bar{G}}^2 - a_k\|x_k - T(x_k)\|_{\bar{G}}^2 [\frac{\|x_k - x^*\|_{\bar{G}}^2 - \|T(x_k) - T(x^*)\|_{\bar{G}}^2}{\|x_k - T(x_k)\|_{\bar{G}}^2} + (1 - a_k)].$$

That is,

$$\|x_{k+1} - x^*\|_{\bar{G}}^2 = \|x_k - x^*\|_{\bar{G}}^2 - a_k A_k(x^*)\|x_k - T(x_k)\|_{\bar{G}}^2. \tag{7}$$

Relation (7) and assumption A1 imply that the sequence $\|x_k - x^*\|_{\bar{G}}^2$ is nonincreasing and, therefore, is convergent. This result implies that either (i) $\|x_k - T(x_k)\|_{\bar{G}}^2$ converges to zero and, therefore, Proposition 4 implies that every limit point of the sequence of iterates $\{x_k\}$ is also a fixed point solution, or (ii) $a_k A_k(x^*)$ converges to zero and then assumption A2 implies that every limit point of the sequence of iterates $\{x_k\}$ is a fixed point solution. If either (i) or (ii) holds, then the entire sequence of iterates $\{x_k\}$ converges to a fixed point solution since as we already established in (7), $\|x_k - x^*\|_{\bar{G}}^2$ is convergent for every fixed point x^*. Q.E.D.

This result extends Banach's fixed point theorem since when T is a contraction, a choice of $a_k = 1$ for all k satisfies assumptions A1 and A2 (see also Lemma 2).

As we next show, this theorem also includes as special case Dunn's averaging results (see [13], [38]).

Theorem 2 : *Suppose that T is a nonexpansive map and the step sizes a_k from the class C are a Dunn sequence and that T has a fixed point solution x^*. Then the map T and the step sizes a_k in the class C satisfy assumptions A1 and A2.*

Proof: Assume as in Dunn's Theorem that T is a nonexpansive map and that $a_k \in [0,1]$ satisfies the condition $\sum_k a_k(1-a_k) = +\infty$. Since T is a nonexpansive map and $a_k \in [0,1]$, $A_k(x^*) \geq 0$ (that is, A1 holds).

To establish the validity of assumption A2, we suppose that it does not hold. If $a_k A_k(x^*)$ converges to zero, some limit point of the sequence $\{x_k\}$ is not a solution. In the proof of Theorem 1, we showed that

$$\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 \leq -a_k A_k(x^*)\|x_k - T(x_k)\|,$$

which implies (since $a_k A_k(x^*) \geq 0$) that the sequence $\{\|x_k - x^*\|^2\}$ converges for every fixed point solution x^*. Corollary 4 implies that since $a_k \in [0,1]$, the sequence $\{\|x_k - T(x_k)\|\}$ converges. Since we assumed that some limit point of the sequence $\{x_k\}$ will not be a solution, $\|x_k - T(x_k)\| \geq B > 0$ for all $k \geq k_0$, for sufficiently large constant k_0. Then for $k \geq k_0$,

$$\lim_k \|x_{k+1} - x^*\|^2 - \|x_{k_0} - x^*\|^2 \leq -\sum_{k=k_0} a_k A_k(x^*)B.$$

But since T is nonexpansive, $\sum_k a_k A_k(x^*) \geq \sum_k a_k(1-a_k) = +\infty$. Therefore, $\sum_k a_k A_k(x^*) = +\infty$.

This result is a contradiction since it implies that

$$+\infty > \lim_k \|x_{k+1} - x^*\|^2 - \|x_{k_0} - x^*\|^2 \leq -\sum_{k=k_0} a_k A_k(x^*)B = -\infty.$$

Therefore, if $a_k A_k(x^*)$ converges to zero, then every limit point of the sequence $\{x_k\}$ is a fixed point solution (that is, assumption A2 is valid). We conclude that the assumptions of Dunn's theorem imply the assumptions of Theorem 1 and, therefore, Dunn's theorem becomes a special case. Q.E.D.

Example:

Consider the map $T(x) = x\sqrt{1 - \|x\|}$ over the set $\{x : \ \|x\| \leq 1\}$. The fixed point solution $x^* = 0$ is unique. The map T is nonexpansive around solutions, since

$$\|T(x) - T(x^*)\| = \|x\|\sqrt{1 - \|x\|} \leq \|x\| = \|x - x^*\|.$$

If we choose step sizes $a_k = 1$ for all k, then Dunn's averaging result does not apply since $\sum_k a_k(1 - a_k) = 0 < +\infty$. Banach's fixed point theorem also does not apply since T is a nonexpansive but not a contractive map. Nevertheless, Theorem 1 ensures convergence since $a_k = 1$ for all k is bounded away from zero and $A_k(x^*) = \frac{\|x_k - x^*\|^2 - \|T(x_k) - x^*\|^2}{\|x_k - T(x_k)\|^2} + 1 - a_k = \frac{\|x_k\|}{\|1 - \sqrt{1 - \|x_k\|}\|^2}$ and therefore, assumptions A1 and A2 hold. Therefore, Theorem 1 applies for this choice of the map T and step sizes.

Remarks:

1. Observe that the iterates we considered in Theorem 1 do not necessarily require that the step sizes lie between zero and one.

2. Theorem 1 does not require the map T to be nonexpansive. Rather it requires assumptions A1 and A2. The convergence result depends upon not only the step sizes a_k, but also the quantity $A_k(x^*) = \frac{\|x_k - x^*\|_{\bar{G}}^2 - \|T(x_k) - T(x^*)\|_{\bar{G}}^2}{\|x_k - T(x_k)\|_{\bar{G}}^2} + (1 - a_k)$ which measures both how far the step size a_k lies from one and "how far" the map T lies from "nonexpansiveness" relative to some $\bar{G}$ norm. Next we examine how restrictive assumptions A1 and A2 are and how we can replace them with the various versions of coerciveness we defined in Definition 1.

   (a) If $S \subseteq [0, 1]$ and T is a *nonexpansive* map relative to a $\bar{G}$ norm around fixed point solutions, then assumptions A1 and A2 are valid, that is,

   $$a_k A_k(x^*) = a_k \frac{\|x_k - x^*\|_{\bar{G}}^2 - \|T(x_k) - T(x^*)\|_{\bar{G}}^2}{\|x_k - T(x_k)\|_{\bar{G}}^2} + a_k(1 - a_k) \geq a_k(1 - a_k) \geq 0. \quad (8)$$

   Conclusion from (a):

   If T is a *nonexpansive* map relative to a $\bar{G}$ norm and the set $S \subseteq [0, 1]$, then all schemes that "repel" a_k from zero and one, unless at a solution, induce a sequence of iterates that converges to a solution.

   (b) As a special case of subcase (a) assume that the step size search set is $S = [c_1, c_2]$ for some $0 < c_1 < c_2 < 1$ (see, for example, scheme III in the next section) and that T is a nonexpansive map. Relation (8) becomes $a_k A_k(x^*) \geq c_1(1 - c_2) > 0$. Then assumptions A1 and A2 follow easily. This observation is valid regardless of the choice of potential g.

(c) Theorem 1 is valid for maps T that are weaker than nonexpansive. What happens if a map T satisfies the *nonexpansive-contractiveness* condition around solutions x^*? Then for step sizes $0 \leq a_k \leq c < 1 - B$, $A_k(x^*) \geq 1 - B - a_k \geq 1 - B - c > 0$. Assumptions A1 and A2 follow for all schemes that "repel" a_k from zero unless at a solution. It is important to observe that there are maps T satisfying the *nonexpansive-contractiveness* condition that are weaker than nonexpansive.

Conclusion from (c):

If $S = [0, c] \subseteq [0, 1-B)$ and T is a *nonexpansive-contractive* map with coercivity constant $0 \leq B < 1$, then schemes that "repel" a_k from zero unless at a solution induce a sequence of iterates that converges to a solution.

(d) Another class of maps T weaker than nonexpansive that satisfy Theorem 1 are maps that satisfy the following condition

$$\|T(x) - T(x^*)\|_{\bar{G}}^2 \geq \|x - x^*\|_{\bar{G}}^2 + B\|x - T(x)\|_{\bar{G}}^2 \quad \forall x \in K, \text{ for some constant } B > 1. \quad (9)$$

Maps T satisfying this condition are expansive. Then for step sizes $0 \geq a_k \geq c > 1 - B$, $A_k(x^*) \leq 1 - B - a_k \leq 1 - B - c < 0$. Assumptions A1 and A2 follow for all schemes that "repel" a_k from zero unless at a solution.

Conclusion from (d):

If $S = [c, 0] \subseteq (1 - B, 0]$ and T is a map satisfying condition (9), then schemes that "repel" a_k from zero unless at a solution induce a sequence of iterates that converges to a solution.

(e) Finally, we observe that for contractive maps, if we consider step sizes in $S = [0.1]$ and schemes that "repel" a_k from zero unless at a solution, then assumptions A1 and A2 are also valid. The following lemma illustrates this observation.

Lemma 2 : *Let T be a contractive map relative to a $\|.\|_G$ norm. Then for any choice of step sizes $a_k \in [0, 1]$ and schemes that "repel" a_k from zero unless at a solution, the sequence $\{x_k\}$ satisfies assumptions A1 and A2.*

Proof: If T is a contractive map relative to the $\bar{G}$ norm with a contractive constant $A \in (0, 1)$, then Corollary 2 implies that

$$A_k(x^*) = \frac{2(F(x_k) - F(x^*))^t \bar{G}(x_k - x^*)}{\|F(x_k) - F(x^*)\|_{\bar{G}}^2} - a_k \geq (1 - A)\frac{\|x_k - x^*\|_{\bar{G}}^2}{\|F(x_k) - F(x^*)\|_{\bar{G}}^2} + 1 - a_k \geq \quad (10)$$

$$(1-A)\frac{\|x_k - x^*\|_G^2}{\|F(x_k) - F(x^*)\|_G^2} \geq 0.$$

It is easy to see that expression (10) and the fact that the sequence $\|x_k - x^*\|_G^2$ is nonincreasing implies assumptions A1 and A2. Q.E.D.

Conclusion from (e):

If $S = [0,1]$ and T is a contractive map, then algorithmic schemes that "repel" a_k from zero unless at a solution, induce a sequence of iterates that converge to a solution.

The discussion in (a)-(e) suggests that in order to satisfy assumptions A1 and A2 in Theorem 1, we may either (i) restrict the line searches we perform to a set $S \subseteq [0,1)$ or $(-1,0]$, so that our results apply for maps that satisfy a condition weaker than nonexpansiveness, or (ii) extend the line searches to a set $S \supseteq [0,1]$ or $[-1,0]$ but as a result we might need to impose stronger assumptions either on the map T or on the step sizes a_k.

Generalized Norms

Our analysis so far applies if we consider the conditions of contractiveness, nonexpansiveness and strictly weak nonexpansiveness relative to a $\|.\|_G$ norm. We observe that the analysis also applies if we impose versions of the conditions with respect to a generalized norm P.

In particular, we might consider a generalized norm $P : K \times K \to R$ satisfying the following properties:

1. $P(x^*, T(x^*)) = 0$ if and only if x^* is a solution.

2. $P(x, x) = 0$.

3. $P(x, T(x)) \geq 0$, $P(x, x^*) \geq 0$.

4. P is *convex* relative to the first component, that is, $P(y, .)$ is convex for all fixed y, and for all points x_1, z_1 and $a \in R$, some constant $D \geq 0$ satisfies the condition

$$P(ax_1 + (1-a)z_1, y) \leq aP(x_1, y) + (1-a)P(z_1, y) - D_1 a(1-a)P(z_1, x_1).$$

(If $D_1 > 0$, then P is *strictly convex*).

Some Examples of Generalized Norms:

1) $P(x, y) = \|x - y\|_G$ or $P(x, y) = \|x - y\|_G^2$ for some positive definite and symmetric matrix G.

2) For variational inequality problems $VI(f,K)$, let $P(x,y) = f(x)^t(x-y)$. Suppose $f(x) = Mx - c$ and the matrix M is positive semidefinite. Then a possible choice of a map T could be $T(x) \in argmin_{y \in K} f(x)^t y$.
3) For variational inequality problems $VI(f,K)$, let $P(x,y) = (f(x) - f(y))^t(x-y)$.

Using generalized norms, we can extend the notions of coerciveness as follows.

Definition 4 : *A map T is* **coercive** *around solutions x^* relative to a generalized norm P, if for some constant $A \geq 0$ and for all x*

$$P(T(x), x^*) \leq \sqrt{A} P(x, x^*).$$

If $A = 1$, then T is a **nonexpansive** *map relative to the generalized norm P around solutions. If $A < 1$, then T is a* **contractive** *map relative to the generalized norm P around solutions.*

It will be useful in our subsequent analysis to observe that if $P(x,x^*) = 0$ and $P(T(x), x^*) = 0$ then $P(x, T(x)) = 0$.
This observation follows from property 4 of the generalized norm conditions. Property 1 then also implies that x is a solution.

Let us now consider a generalized norm P with the following two additional assumptions.

5. . The generalized norm P is *convex* with respect to the second component. That is, for a fixed x and for all points z_2, y_2 and $a \in R^+$, some constant $D_2 \geq 0$, satisfying

$$P(x, ay_2 + (1-a)z_2) \leq aP(x, y_2) + (1-a)P(x, z_2) - D_2 a(1-a)P(z_2, y_2).$$

(If $D_2 > 0$, then P is *strictly convex*).

6. The generalized norm P satisfies the triangle inequality,

$$P(x,y) \leq P(x,z) + P(z.y).$$

Lemma 3 : *Under assumptions 1-6 on the generalized norm P and for a map T that is coercive relative to P, for all $a \in [0.1]$,*

$$P(x_k(a), T(x_k(a))) \leq [1 - a(1-\sqrt{A})] P(x_k, T(x_k)).$$

Proof: The convexity assumption 4 implies that for all $a \in [0,1]$,

$$P(x_k(a), T(x_k(a))) \leq aP(T(x_k), T(x_k(a))) + (1-a)P(x_k, T(x_k(a)))$$

(then the triangle inequality 6 implies that)

$$\leq P(T(x_k), T(x_k(a))) + (1-a)P(x_k, T(x_k))$$

(the definition of coerciveness further implies that)

$$\leq \sqrt{A}P(x_k, x_k(a)) + (1-a)P(x_k, T(x_k))$$

(the convexity assumption 5 implies that)

$$\leq a\sqrt{A}P(x_k, T(x_k)) + (1-a)P(x_k, x_k) + (1-a)P(x_k, T(x_k))$$

(finally, assumption 2 implies that)

$$\leq a\sqrt{A}P(x_k, T(x_k)) + (1-a)P(x_k, T(x_k)).$$

Q.E.D.

In the previous analysis, we presented choices of maps T and step sizes a_k that imply assumptions A1 and A2. Next we illustrate a specific class of potentials satisfying assumptions A1 and A2 as well. This class of potentials is only a special case. Theorem 1 also applies to other potential functions as we will show in detail in Section 3. To state these results in a more general form, we use the generalized norm concept that we have introduced in this section.

2.2 A Class of Potential Functions

The previous discussion related assumptions A1 and A2 to the notions of contractiveness, nonexpansiveness, and strictly weak nonexpansiveness. We next address the following natural question: *Is there a class of potentials satisfying assumptions A1 and A2?*

To provide partial answers to these questions, we consider potentials of the form

$$g(x_k(a)) = P(x_k(a), T(x_k(a))^2 - \beta h(a)P(x_k, T(x_k))^2.$$

As we will see, these potentials satisfy assumptions A1 and A2 if we impose the following conditions: $P : K \times K \rightarrow R$ is a continuous function, $h : R \rightarrow R^+$ is a given continuous function, and $\beta > 0$ is a given constant satisfying the following assumptions

(i) P is a generalized norm (that is, satisfies assumptions 1-6 from the last section).

(ii) $h(0) = 0$.

(iii) For some point $\bar{a} \in S \cap (0,1)$, $h(\bar{a}) > 0$.

(iv) h is bounded from above on $S \cap [0,1]$, that is, $0 < C = sup_{a \in S \cap [0,1]} h(a) < +\infty$.

(v) $1 < C\beta$.

The following result provides a bound on the rate of convergence of the adaptive averaging framework for choices of potentials satisfying properties (i)-(v).

Proposition 5 *: Consider the class of potentials g satisfying the properties (i)-(iv). The general scheme converges at a rate*

$$P(x_{k+1}, T(x_{k+1}))^2 \leq [1 - \beta(C - h(a_k))] P(x_k, T(x_k))^2. \tag{11}$$

Furthermore, for step sizes $a_k \in [0,1] \cap S$,

$$P(x_{k+1}, T(x_{k+1}))^2 \leq [1 - a_k(1 - \sqrt{A})]^2 P(x_k, T(x_k))^2. \tag{12}$$

Proof: Select $a^* \in S \cap [0,1]$. The iteration of the general averaging framework implies that

$$g(x_{k+1}) = P(x_{k+1}, T(x_{k+1}))^2 - \beta h(a_k) P(x_k, T(x_k))^2 \leq$$

$$P(x_k(a^*), T(x_k(a^*)))^2 - \beta h(a^*) P(x_k, T(x_k))^2.$$

Lemma 3 implies that $P(x_k(a^*), T(x_k(a^*))^2 \leq P(x_k, T(x_k))^2$ and so

$$g(x_{k+1}) = P(x_{k+1}, T(x_{k+1}))^2 - \beta h(a_k) P(x_k, T(x_k))^2 \leq P(x_k, T(x_k))^2 - \beta h(a^*) P(x_k, T(x_k))^2,$$

which, by letting $h(a^*)$ approach C, implies the inequality (11).

On the other hand, if $a_k \in [0,1] \cap S$, then Lemma 3 implies that

$$0 \leq P(x_{k+1}, T(x_{k+1}))^2 \leq (1 - a_k(1 - \sqrt{A}))^2 P(x_k, T(x_k))^2.$$

Q.E.D.

Remarks:

(1) Relation (12) is valid for any choice of β (negative as well as positive).

(2) If C is not a limit point of $h(a_k)$, then for some constant $h < C$, $h(a_k) \leq h$, for all $k \geq k_0$. Therefore, for all $k \geq k_0$,

$$P(x_{k+1}, T(x_{k+1}))^2 \leq [1 - \beta(C - h)] P(x_k, T(x_k))^2. \tag{13}$$

(3) If the step sizes a_k are bounded away from zero, i.e., $a_k \geq c > 0$ for all $k \geq k_0$, then

$$P(x_{k+1}, T(x_{k+1}))^2 \leq [1 - c(1 - \sqrt{A})]^2 P(x_k, T(x_k))^2. \tag{14}$$

The following proposition shows that potentials satisfying properties (i)-(v) also satisfy assumptions A1 and A2.

Proposition 6 : *Consider potentials of the type*

$$g(x_k(a)) = P(x_k(a), T(x_k(a))^2 - \beta h(a) P(x_k, T(x_k))^2.$$

Assume that $P : R^n \times R^n \rightarrow R$ is a continuous function, $h : R \rightarrow R^+$ is a given continuous function, and $\beta > 0$ is a given constant satisfying properties (i)-(v).

Any such potential satisfies the following properties:

a. $\lim_{k \rightarrow +\infty} a_k = 0$ implies that every limit point of $\{x_k\}$ is a fixed point solution (that is, assumption A1 is valid).

b. $h(a_k) \leq 0$ implies that x_k is a fixed point solution.

c. $\lim_{k \rightarrow +\infty} h(a_k) = 0$ implies that every limit point of $\{x_k\}$ is a fixed point solution.

d. If $h(a_k) \leq a_k A_k(x^*)$ and $a_k \geq 0$, then assumptions A1 and A2 are valid.

Proof: Before proving parts *a-d*, we observe that Proposition 5 implies that $\lim_{k \rightarrow +\infty} P(x_k, T(x_k))$ exists.

a. If $\lim_{k \rightarrow +\infty} a_k = 0$, then the iteration of the general scheme implies that

$$\lim_{k \rightarrow +\infty} g(x_{k+1}) = \lim_{k \rightarrow +\infty} P(x_{k+1}, T(x_{k+1}))^2 - \beta h(a_k) P(x_k, T(x_k))^2 =$$

$$\lim_{k \rightarrow +\infty} P(x_k, T(x_k))^2 [1 - \beta h(a_k)] = \lim_{k \rightarrow +\infty} P(x_k, T(x_k))^2 \leq$$

$$\lim_{k \rightarrow +\infty} [P(x_k(a). T(x_k(a))^2 - \beta h(a) P(x_k, T(x_k))^2],$$

for any a. Therefore,

$$\lim_{k \rightarrow +\infty} P(x_k(a), T(x_k(a))^2 \geq (1 + \beta h(a)) \lim_{k \rightarrow +\infty} P(x_k, T(x_k))^2.$$

Select $\bar{a} \in (0, 1] \cap S$ satisfying $h(\bar{a}) > 0$. Then Proposition 5 implies that $P(x_k(\bar{a}), T(x_k(\bar{a}))^2 \leq [1 - \bar{a}(1 - \sqrt{A})]^2 P(x_k, T(x_k))^2$. Therefore,

$$[1 - \bar{a}(1 - \sqrt{A})]^2 \lim_{k \rightarrow +\infty} P(x_k, T(x_k))^2 \geq (1 + \beta h(\bar{a})) \lim_{k \rightarrow +\infty} P(x_k, T(x_k))^2.$$

Consequently, either $\lim_{k\to+\infty} P(x_k, T(x_k)) = 0$, implying (by property (i) and the continuity of P) that every limit point of $\{x_k\}$ is a fixed point solution,

$$[1 - \bar{a}(1-\sqrt{A})]^2 \geq 1 + \beta h(\bar{a}).$$

But $h(\bar{a}) > 0$ implies that $[1 - \bar{a}(1-\sqrt{A})]^2 > 1$ which is a contradiction.

We conclude that for this class of potentials, assumption A1 is valid.

b. Select an $\bar{a}$ so that $\frac{1}{\beta} < h(\bar{a})$. Since $\frac{1}{C} < \beta$, such an $\bar{a}$ exists.

If x_k is not a fixed point solution, then $P(x_k, T(x_k)) \neq 0$. The generalized nonexpansive property and the choice of $\bar{a}$ imply that

$$g(x_k(\bar{a})) = P(x_k(\bar{a}), T(x_k(\bar{a}))^2 - \beta h(\bar{a}) P(x_k, T(x_k))^2 < P(x_k, T(x_k))^2 - P(x_k, T(x_k))^2 = 0.$$

But if $h(a_k) \leq 0$, then

$$g(x_{k+1}) = P(x_{k+1}, T(x_{k+1}))^2 - \beta h(a_k) P(x_k, T(x_k))^2 \geq 0.$$

Therefore, $0 \leq g(x_{k+1}) \leq g(x_k(\bar{a})) < 0$, which is a contradiction.

c. Suppose that $\lim_{k\to+\infty} h(a_k) = 0$ and that $\lim_{k\to+\infty} P(x_k, T(x_k)) \neq 0$. Therefore, as we have shown in part a,

$$\lim_{k\to+\infty} g(x_{k+1}) = \lim_{k\to+\infty} P(x_{k+1}, T(x_{k+1}))^2 = \lim_{k\to+\infty} P(x_k, T(x_k))^2 > 0.$$

The condition $1 - \beta C < 0$ implies that for some $a \in [0,1] \cap S$, $1 - \beta h(a) < 0$. Consequently, since Lemma 3 implies that $P(x_k(a), T(x_k(a)) \leq P(x_k, T(x_k))$,

$$\lim_{k\to+\infty} g(x_k(a)) = \lim_{k\to+\infty} P(x_k(a), T(x_k(a))^2 - \beta h(a) \lim_{k\to+\infty} P(x_k, T(x_k))^2 <$$
$$\lim_{k\to+\infty} P(x_k, T(x_k))^2 - \lim_{k\to+\infty} P(x_k, T(x_k))^2 = 0.$$

Therefore, $0 < \lim_{k\to+\infty} g(x_{k+1}) \leq \lim_{k\to+\infty} g(x_k(a)) < 0$, which is a contradiction. This result implies that $\lim P(x_k, T(x_k)) = 0$ and, therefore, that every limit point is a fixed point solution.

d. Part b and the fact that $h(0) = 0$ imply that $h(a_k) > 0$ and $a_k > 0$ whenever x_k is not a solution. Consequently, if x_k is not a solution, then $A_k(x^*) > 0$ as well (that is, assumption A2 is valid). Moreover, if $\lim_{k\to+\infty} A_k(x^*) = 0$, then $\lim_{k\to+\infty} h(a_k) = 0$. Then part c implies that every limit point of $\{x_k\}$ is a fixed point solution (that is, assumption A2 is valid). Q.E.D.

Remark:

For example, if $h(a) = a(1-a)$ and $\beta > 4$, then we can choose any $\frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{\beta}} < a < \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{\beta}}$ as $\bar{a}$ so that $\frac{1}{\beta} < h(\bar{a})$.

3 Potential Optimizing Methods: Specific Cases

In this section we apply the results from the previous section to special cases. We consider several choices for the potential function g and examine the convergence behavior for the sequence they generate.

Table I summarizes some of these results.

Let $\beta \in R$ be a given constant and

$$g(x_k(a)) = \|x_k(a) - T(x_k(a))\|_G^2 + \beta\|x_k(a) - x_k(0)\|_G^2. \tag{15}$$

We consider several choices for the constant β.

Scheme I: Let $S = [0,1]$ and $\beta \geq 0$, then (15) becomes $g_1(x_k(a)) = \|x_k(a) - T(x_k(a))\|_G^2 + \beta\|x_k(a) - x_k(0)\|_G^2$.

This potential stems from (5) and the observation that we do not allow the iterates to lie "far away" from each other.

Lemma 4 : *Suppose T is a contraction relative to the $\|.\|_G$ norm and $\beta \leq 1 - A$. Then if $\lim_{k\to+\infty} a_k = 0$, the sequence $\{x_k\}$ converges to a solution (that is, assumption A1 is valid).*

Proof: If $\lim_{k\to+\infty} a_k = 0$, then

$$\lim_{k\to+\infty} \|x_k - T(x_k)\|_G^2 \leq \lim_{k\to+\infty} \|T(x_k) - T^2(x_k)\|_G^2 + \beta \lim_{k\to+\infty} \|x_k - T(x_k)\|_G^2 \leq$$

(since $\beta < 1 - A$)

$$(A + \beta) \lim_{k\to+\infty} \|x_k - T(x_k)\|_G^2 < \lim_{k\to+\infty} \|x_k - T(x_k)\|_G^2$$

Therefore, $\lim_{k\to+\infty} \|x_k - T(x_k)\|_G^2 = 0$. Proposition 4 and Corollary 4 imply that the sequence $\{x_k\}$ converges to a fixed point solution. Q.E.D.

Theorem 3 : *If T is a contractive map relative to the $\|.\|_G$ norm, with a contraction constant $0 < A < 1$, then the sequence induced by scheme I converges to a solution for any choice $0 \leq \beta < 1 - A$.*

Proof:

- First, we restrict the line searches so that $a_k \in [0,1]$.

In this case, the proof follows from Theorem 1. In fact, Lemma 4 implies that for any choice $\beta < 1 - A$, the step sizes a_k satisfy assumption A1. Moreover, Lemma 2 implies that $A_k(x^*)$ satisfies assumptions A1 and A2.

- More generally we perform unrestricted line searches in the sense of choosing step sizes $a_k \in R^+$.

Then the proof follows since the general iteration of the algorithm implies that

$$\|x_{k+1} - T(x_{k+1})\|_G^2 + \beta a_k^2 \|x_k - T(x_k)\|_G^2 \leq \|x_k - T(x_k)\|_G^2.$$

Then

$$\|x_{k+1} - T(x_{k+1})\|_G^2 - \|x_k - T(x_k)\|_G^2 \leq -\beta a_k^2 \|x_k - T(x_k)\|_G^2 \leq 0.$$

Therefore,

$$\lim_{k \to +\infty} a_k^2 \|x_k - T(x_k)\|_G^2 = 0.$$

This result implies that (i) either $\lim_{k \to +\infty} a_k = 0$, or (ii) $\lim_{k \to +\infty} \|x_k - T(x_k)\|_G^2 = 0$. In the former case, Lemma 4 implies that for any choice $\beta < 1 - A$, the sequence $\{x_k\}$ converges to a solution. In case (ii), Proposition 3 implies that the sequence $\{x_k\}$ converges to a solution. Q.E.D.

As the next corollary shows, scheme I extends to also include *nonexpansive* affine maps relative to the $\|.\|_G$ norm.

Corollary 5 *: If T is an affine, nonexpansive map, then the iterates $\{x_k\}$ of scheme I converge to a fixed point solution.*

Proof: Let $d_k = T(x_k) - x_k$. For *affine* problems, the map $M = I - T$ is a positive semidefinite matrix, and since $Mx_k = -d_k$,

$$a_k = \frac{d_k^t M^t G d_k}{\|Md_k\|_G^2 + \beta \|d_k\|_G^2}. \tag{16}$$

Observe that if we let $\bar{G} = M^t GM + \beta G$ in the definition of $A_k(x^*)$, then replacing this choice of $\bar{G}$ and accounting for the fact that $M^t G$ is positive semidefinite, we see that

$$A_k(x^*) \geq \frac{d_k^t M^t G d_k + \beta (x_k - x^*)^t M^t G (x_k - x^*)}{\|Md_k\|_G^2 + \beta \|d_k\|_G^2} \geq \beta \frac{(F(x_k) - F(x^*))^t G (x_k - x^*)}{\|Md_k\|_G^2 + \beta \|d_k\|_G^2}.$$

Strong-f-monotonicity of f (or, equivalently, nonexpansiveness of T) easily implies that $A_k(x^*) \geq 0$ and that if either a_k or $A_k(x^*)$ converge to zero, then $\{x_k\}$ converges to a solution (that is, assumptions A1 and A2 hold). Theorem 1 implies the conclusion. Q.E.D.

Lemma 5 *: If T is affine and is domain and range coercive with coercivity constants A and B, then all iterates with $d_k \neq 0$ satisfy the inequality $a_k \geq \frac{1-B}{2} + \frac{1-A-\beta(1-B)}{2(1+\sqrt{A})^2+\beta)}$. If T is nonexpansive, that is $0 < A \leq 1$ and $B = 0$, then for any choice of $\beta < 1 - A$, a_k is bounded away from $1/2$.*

Proof: Lemma 1 implies, since T is domain and range coercive with coercivity constants A and B, that $w^t GMw \geq \frac{1-A}{2} \|w\|_G^2 + \frac{1-B}{2} \|Mw\|_G^2$. Moreover, if $d_k \neq 0$, then Proposition 3 implies that

$$a_k = \frac{d_k^t M^t G d_k}{\|Md_k\|_G^2 + \beta \|d_k\|_G^2} \geq \frac{1-B}{2} + \frac{1 - A - \beta(1-B)}{2(1+\sqrt{A})^2 + \beta)}. \tag{17}$$

Q.E.D.

The following proposition provides a rate of convergence for scheme I.

Proposition 7 : *For maps T that are contractive relative to the G norm with a contraction constant A satisfying the condition $\beta < 1 - A$, the iterates generated by scheme I satisfy the estimates:*

$$\|x_{k+1} - T(x_{k+1})\|_G^2 \leq [A + \beta(1 - a_k^2)]\|x_k - T(x_k)\|_G^2. \tag{18}$$

Proof: The iteration of scheme I implies the inequality

$$\|x_{k+1} - T(x_{k+1})\|_G^2 + \beta a_k^2\|x_k - T(x_k)\|_G^2 \leq \|T(x_k) - T^2(x_k)\|_G^2 + \beta\|x_k - T(x_k)\|_G^2.$$

Since the map T is a contraction,

$$\|x_{k+1} - T(x_{k+1})\|_G^2 \leq (A + \beta(1 - a_k^2))\|x_k - T(x_k)\|_G^2,$$

implying (18). If $\beta < 1 - A$, then $I - T$ is a contraction with coercivity constant $0 < A + \beta < 1$. Q.E.D.

Remark:

If $a_k >> 1$, then scheme I has a strictly better rate of convergence than the classical iteration (2).

Observe when T is an affine, firmly contractive map, then a choice of $0 < \beta << \frac{1-A}{2}$ and Lemma 5 imply that

$$a_k = \frac{d_k^t M^t G d_k}{\|Md_k\|_G^2 + \beta\|d_k\|_G^2} >> 1.$$

Requiring the iterates to lie close to each other might be too much of a restriction. In fact, Proposition 7 shows that the rate of convergence of scheme I is, in the nonlinear contractive case, worse than the rate of convergence of the classic iteration (2) unless the step size a_k is greater than 1. Therefore, motivated as before by the minimization problem (5), we consider the following potential.

Scheme II: Let $\beta = 0$ in expression (15). Then $g_2(x_k(a)) = \|x_k(a) - T(x_k(a))\|_G^2$ and $S = R^+$. Itoh, Fukushima and Ibaraki [28] have studied a line search scheme of this type in the context of unconstrained variational inequality problems. They consider only strongly monotone problem functions, which correspond to contractive fixed point maps.

Remark:

Let $T(x) = x - Mx + c$ be an *affine, nonexpansive* map relative to the $\|.\|_G$ norm and let $d_k = T(x_k) - x_k$. Then $a_k = argmin_{\{a \geq 0\}}\ g_2(x_k(a)) = argmin_{\{a \geq 0\}}\|Mx_k(a) - c\|_G^2 = \frac{d_k^t GM(d_k)}{\|Md_k\|_G^2}$.

Lemma 6 *: If T is an affine, contractive map relative to the $\|.\|_G$ norm, with coercivity constant $0 \leq A \leq 1$, then for all iterates k with $d_k \neq 0$, $a_k \geq \frac{1}{2} + \frac{1-\sqrt{A}}{2(1+\sqrt{A})}$. Therefore, if $A < 1$, then a_k is bounded away from $\frac{1}{2}$.*

Proof: The proof follows from Lemma 5 with $\beta = B = 0$.

These results show that $a_k \geq \frac{1}{2}$ in the nonexpansive case (when $A = 1$).

Lemma 7 *: If T is a contraction with coercivity constant A, then $\lim_{k\to\infty} a_k = 0$ implies that every limit point of the sequence of iterates $\{x_k\}$ is a fixed point solution (that is, assumption A1 holds).*

Proof: If $\lim_{k\to\infty} a_k = 0$, but $\lim_{k\to\infty} \|x_k - T(x_k)\|_G \neq 0$. Then,

$$\|x_k - T(x_k)\|_G^2 \leq \|T(x_k) - T(T(x_k))\|_G^2.$$

Since T is a contraction mapping, $\|T(x_k) - T(T(x_k))\|_G^2 \leq A\|x_k - T(x_k)\|_G^2$ and, therefore, $\lim_{k\to\infty} \|x_k - T(x_k)\|_G^2 \leq A \lim_{k\to\infty} \|x_k - T(x_k)\|_G^2$, which is a contradiction. Consequently $\lim_{k\to\infty} \|x_k - T(x_k)\|_G = 0$ and so Proposition 4 implies the conclusion. Q.E.D.

Remark:

The proof of this lemma also follows from Proposition 6 part a with $\beta = 0$. The class of potential described in this proposition need to satisfy properties (i)-(v), and in particular, to be contractive with constant $0 < A < 1$.

The following theorem shows when scheme II works.

Theorem 4 *: If T is an affine, nonexpansive map relative to the $\|.\|_G$ norm, or if T is a nonlinear, contractive map relative to the $\|.\|_G$ norm with coercivity constant A, then the sequence that scheme II induces converges to a fixed point solution.*

Proof: First, we illustrate how the proof of this theorem follows from Theorem 1. Lemma 7 implies that the step sizes a_k satisfy assumption A1. Moreover,

1) If T is a *contractive* map relative to the $\|.\|_G$ norm and our linesearch is restricted in $[0, 1]$, then Lemma 2 implies assumptions A1 and A2.

2) In the *affine, nonexpansive* case, we do not need to restrict the linesearch. In fact,

$$a_k = \frac{d_k^t GM(d_k)}{\|Md_k\|_G^2} \geq \frac{1}{2}$$

and so a choice of $\bar{G} = M^t GM$ implies, using strong-f-monotonicity, that $A_k(x^*) = a_k \geq \frac{1}{2}$. That is, assumptions A1, A1 and A2 are valid.

In both cases, Theorem 1 implies the result.

Alternatively, for nonlinear, contractive mappings the result follows from the observation that

$$\|x_{k+1} - T(x_{k+1})\|_G^2 \leq \|T(x_k) - T(T(x_k))\|_G^2 \leq A\|x_k - T(x_k)\|_G^2.$$

This result implies that the sequence $\{\|x_k - T(x_k)\|_G^2\}$ converges to zero. Proposition 3 then implies the conclusion. Q.E.D.

The following proposition provides a rate of convergence for scheme II.

Proposition 8 *: For nonlinear, contractive maps relative to the* $\|.\|_G$ *norm, with contraction constant* $A \in (0,1)$*, the line search we considered in scheme II gives rise to the estimates:*

$$\|x_{k+1} - T(x_{k+1})\|_G^2 \leq A\|x_k - T(x_k)\|_G^2. \tag{19}$$

Proof: For nonlinear, contractive mappings T,

$$\|x_{k+1} - T(x_{k+1})\|_G^2 \leq \|T(x_k) - T(T(x_k))\|_G^2 \leq A\|T(x_k) - x_k\|_G^2.$$

Q.E.D.

In the following discussion, to improve on the rate of convergence of the general scheme, we consider potentials that involve a penalty term that pulls the iterates away from zero unless they are approaching a solution.

Scheme III: Let $\beta \leq 0$ in expression (15). This is equivalent to replacing β with $-\beta$ in (15) and letting $\beta \geq 0$. Then $g_3(x_k(a)) = \|x_k(a) - T(x_k(a))\|_G^2 - \beta\|x_k(a) - x_k(0)\|_G^2$. Moreover, we set $S = [0, c_1]$, with $0 < c_1 \leq 1$.

Alternatively, letting $P(x, T(x)) = \|x - T(x)\|_G$ and $h(a) = a^2$ and $\beta \geq 0$,

$$g_3(x_k(a)) = P(x_k(a), T(x_k(a)))^2 - \beta h(a) P(x_k, T(x_k))^2.$$

Remarks:

1) For *affine, nonexpansive* maps relative to the $\|.\|_G$ norm, we compute

$$a_k = argmin_{\{a \in [0,c_1]\}} g_3(x_k(a)).$$

Let $\bar{a}_k = \frac{d_k^t G(M) d_k}{\|Md_k\|_G^2 - \beta\|d_k\|_G^2}$,

$$\text{if} \quad \|Md_k\|_G^2 \leq \beta\|d_k\|_G^2 \quad \text{or} \quad \bar{a}_k > c_1 \quad \text{then} \quad a_k = c_1, \quad \text{otherwise,} \quad a_k = \bar{a}_k.$$

2) Observe that when T is a contractive map with a contractive constant $A < 1$. Proposition 2 implies that for a choice of $\beta < (\frac{1-A}{2})^2$, $\|Md_k\|_G^2 > \beta\|d_k\|_G^2$.

Lemma 8 : *Let T be an affine, contractive map, with a contractive constant $0 < A \leq 1$. Then if $\|Md_k\|_G^2 > \beta\|d_k\|_G^2$ and $\beta < 4$,*

$$\bar{a}_k \geq \frac{1}{2} + \frac{\beta + 1 - A}{2((1+\sqrt{A})^2 - \beta)}.$$

Proof: The proof follows from Lemma 5 with $-\beta$ replacing β and $B = 0$. Q.E.D.
In the nonexpansive case, $A = 1$. Therefore, that $a_k \geq \frac{1}{2} + \frac{\beta}{2((1+\sqrt{A})^2-\beta)}$.

Lemma 9 : *If T is a nonexpansive map relative to the $\|.\|_G$ norm, then A1 holds.*

Proof: Observe that in scheme III, $h(a) = a^2$, $a \in [0, c_1]$. Consequently, this lemma becomes a special case of part a in Proposition 6. Q.E.D.

Theorem 5 : *If T is a nonexpansive map relative to the $\|.\|_G$ norm with a coercivity constant $0 < A \leq 1$ and $a_k \leq c_1 \leq 1$, then the sequence induced by scheme III converges to a fixed point solution.*

Proof: First, let $c_1 < 1$. The facts that $a_k \leq c_1 < 1$ and T is a nonexpansive map imply that $A_k(x^*) \geq 1 - c_1 > 0$ and, as a result, that assumptions A1 and A2 are valid. Moreover, Lemma 9 implies assumption A2 and Theorem 1 implies the result.

If we let $c_1 = 1$, then we need to assume that T is a contractive map, that is, $0 < A < 1$. Assumptions A1 and A2 are valid because $A_k(x^*) \geq (1 - A)\frac{\|x_k - x^*\|_G^2}{\|F(x_k) - F(x^*)\|_G^2}$. Theorem 1 again implies the result. Q.E.D.
The following proposition provides a rate of convergence for scheme III.

Proposition 9 : *For a contractive (or nonexpansive) map T relative to the G norm with a coercivity constant $0 < A \leq 1$, the line search we consider in scheme III gives rise to estimates:*

$$\|x_{k+1} - T(x_{k+1})\|_G^2 \leq (A - \beta(c_1^2 - a_k^2)))\|x_k - T(x_k)\|_G^2, 0 < c_1 \leq 1. \tag{20}$$

Proof: In scheme III, $P(x, y) = \|x - y\|_G$ and $h(a) = a^2$, $a \in [0, c_1]$, with $0 < c_1 \leq 1$. Then $a^* = c_1$ and $C = c_1^2$. Then as in the proof of Proposition 5, the iteration of scheme III implies that

$$\|x_{k+1} - T(x_{k+1})\|_G^2 - \beta a_k^2\|x_k - T(x_k)\|_G^2 \leq \|x_k(c_1) - T(x_k(c_1))\|_G^2 - \beta c_1^2\|x_k - T(x_k)\|_G^2.$$

Therefore,

$$\|x_{k+1} - T(x_{k+1})\|_G^2 \leq (A - \beta(c_1^2 - a_k^2)))\|x_k - T(x_k)\|_G^2.$$

Q.E.D.

Remarks:

1. Observe that when the map T is contractive, $0 < A < 1$, scheme III is at least as good as the classical iterative method (2). Scheme III has a better rate of convergence than the classical iterative method (2) when, for example, the step sizes a_k converge to zero.

2. The choice of potential in scheme III requires that for nonexpansive maps, we perform a line search with step sizes bounded away from 1. *"How far" should the step sizes we consider lie away from 1?* Perhaps considering step sizes bounded away from 1 is too much of a restriction. For this reason, we modify the potential function of scheme III as follows,

Scheme IV: $g_4(x_k(a)) = \|x_k(a) - T(x_k(a))\|_G^2 - \beta\|x_k(a) - x_k(0)\|_G.\|x_k(a) - x_k(1)\|_G$, and set $S = R^+$.

Proposition 10 *: For nonexpansive maps T relative to the G norm,*

$$\|x_{k+1} - T(x_{k+1})\|_G^2 \leq \|x_k(\frac{1}{2}) - T(x_k(\frac{1}{2}))\|_G^2 - \beta(a_k - \frac{1}{2})^2\|x_k - T(x_k)\|_G^2 \leq \tag{21}$$

$$[1 - \beta(a_k - \frac{1}{2})^2]\|x_k - T(x_k)\|_G^2.$$

Proof: Observe that $P(x,y) = \|x - y\|_G^2$ and $a^* = \frac{1}{2}$. Therefore, the result follows from Proposition 5. Q.E.D.

Remarks:

1. The sequence $\{\|x_k - T(x_k)\|_G\}$ is nonincreasing and, therefore, converges despite the fact that we did not restrict the line search in the set $S = [0, 1]$.

2. Proposition 10 provides a convergence rate for scheme IV.

Lemma 10 *: If T is a nonexpansive map relative to the $\|.\|_G$ norm, then A1 is valid. Moreover, if $\lim_{k \to +\infty} a_k = 1$, then every limit point of the sequence of iterates $\{x_k\}$ is a fixed point solution.*

Proof: The proof follows from parts a and c in Proposition 6. Q.E.D.

The following theorem provides a convergence result as well as a convergence rate for this choice of potential function.

Theorem 6 *: For nonexpansive maps T relative to the G norm, the sequence $\{x_k\}$ that scheme IV generates converges to a fixed point solution.*

Proof: Lemma 10 implies that assumption A1 is valid. Proposition 10 implies that if a limit point of $\{x_k\}$ is not a solution, then a_k converges to $\frac{1}{2}$. This result combined with the nonexpansiveness of the map T, implies that for k_0 large enough, for all $k \geq k_0$, $A_k(x^*) \geq 1 - a_k \geq 0$, that is, assumption A2 applies. The proof then follows from Theorem 1. Q.E.D.

Remark:

For *affine, nonexpansive* maps T relative to the $\|.\|_G$ norm,

$$a_k = \frac{\beta\|d_k\|_G^2 + 2d_k^t GMd_k}{2\beta\|d_k\|_G^2 + 2\|Md_k\|_G^2}.$$

Lemma 11 *: Suppose T is an affine, nonexpansive map relative to the $\|.\|_G$ norm, with a coercivity constant $0 < A \leq 1$. Then*

$$a_k \geq \frac{1}{2} + \frac{1-A}{2\beta + 2(1+\sqrt{A})^2}.$$

If T is a contraction, that is, $0 < A < 1$, then a_k is bounded away from $\frac{1}{2}$.

Proof:

$$a_k \geq \frac{\beta\|d_k\|_G^2 + \|Md_k\|_G^2 + (1-A)\|d_k\|_G^2}{2\beta\|d_k\|_G^2 + 2\|Md_k\|_G^2}.$$

Therefore, whenever $d_k \neq 0$, $a_k \geq \frac{1}{2} + \frac{1-A}{2\beta+2(1+\sqrt{A})^2}$. Q.E.D.

Observe that in the contractive case, a choice of $\beta \geq 1 - (\frac{1-A}{2})^2$ implies that all $a_k \leq 1$.

Corollary 6 *: For affine, nonexpansive maps T relative to the G norm, the sequence $\{x_k\}$ that scheme IV generates converges to a fixed point solution.*

Proof: In the affine case of scheme IV, $a_k = \frac{\beta\|d_k\|_G^2 + 2d_k^t G(M)d_k}{2\beta\|d_k\|_G^2 + 2\|Md_k\|_G^2} \geq \frac{1}{2}$. Then if we choose $\bar{G} = 2M^tGM + 2\beta G$ in the definition of $A_k(x^*)$, then $A_k(x^*) = \frac{2d_k^t M^t G d_k + 3\beta\|d_k\|_G^2}{2\beta\|d_k\|_G^2 + 2\|Md_k\|_G^2} \geq \frac{1}{2}$ since f is strongly-f-monotone relative to the $\|.\|_G$ norm. Therefore, assumptions A1 and A2 are valid and so Theorem 1 implies the result. Q.E.D.

Remarks:

1. The previous analysis does not restrict the line search to the set $S = [0, 1]$. If we do restrict the search to $S = [0, 1]$ then the following theorem implies convergence.

   Theorem 7 *: If T is a nonexpansive map relative to the G norm, then the sequence that scheme IV generates converges to a fixed point solution.*

Proof: The proof follows from Theorem 1. Lemma 10 implies assumption A1. Moreover, assumptions A1 and A2 follow from Lemma 2. Q.E.D.

2. Observe that the previous analysis did not restrict the choice of the constant β. If we choose $\beta > 4$, then the line search in scheme IV always yields a step length $a_k \leq 1$ unless the current point is a solution. This result follows from property (2) of Proposition 6.

3. We can view scheme III as a form of scheme IV if we can find $0 \leq a_4(k) \leq 1$ for which $a_3(k) = \sqrt{a_4(k)(1 - a_4(k))}$. Then

$$x_k(a_3) = x_k(a_4) = x_k + \sqrt{a_4(k)(1 - a_4(k))}(T(x_k) - x_k)$$

and the potential

$$g_3(x_k(a_3)) = g_3(x_k(a_4)) = \|x_k(a_4) - T(x_k(a_4))\|_G^2 - \beta a_4(k)(1 - a_4(k))\|x_k - T(x_k)\|_G^2.$$

Note that we can find these values $a_4(k)$ only if $a_3(k) \leq \frac{1}{2}$. More generally, if $a_3(k) \leq c_1 < 1$, then we can find $0 \leq a_4(k) \leq 1$ and a positive integer m, so that $a_3(k)$ is the mth root of $a_4(k)(1 - a_4(k))$.

4. Suppose a_k is bounded away from $\frac{1}{2}$ by a constant $0 \leq c \neq \frac{1}{2}$. If T is a contractive map, then for a choice of $\beta > \frac{1-A}{(c-\frac{1}{2})^2}$, the rate of convergence for scheme IV is better than that of the classical iterative method (2).

For example, in the contractive case, if the sequence of the step sizes a_k converges either to zero or to a constant $\neq \frac{1}{2}$, then scheme IV strictly improves the rate of convergence of the classical iterative method (2).

Scheme V: $g_5(x_k(a)) = [(x_k(a) - T(x_k(a)))^t(x_k - T(x_k))]^2$, and $S = R^+$.

This method is the classical steepest descent method (see [47]) as studied by Hammond and Magnanti [23], as applied to solving asymmetric system of equations: find $x^* \in K$ satisfying $F(x^*) = x^* - T(x^*) = 0$.

Remark:

If $T(x) = x - Mx + c$ is an *affine* map and $d_k = -(Mx_k - c)$, then

$$min_{a \geq 0}[(x_k(a) - T(x_k(a)))^t(x_k - T(x_k))]^2 = min_{a \geq 0}[(Mx_k(a) - c)^t(Mx_k - c)]^2.$$

implies that

$$a_k = \frac{\|d_k\|^2}{d_k^t M d_k}.$$

Lemma 12 *: If T is a contraction mapping, then $\lim_{k\to+\infty} a_k = 0$ implies that every limit point of the sequence of iterates $\{x_k\}$ is a fixed point solution.*

Proof: Assume T is a contraction map with coercivity constant $A \in (0,1)$. If $\lim_{k\to+\infty} a_k = 0$, then $\lim_{k\to+\infty}[(x_k - T(x_k))^t(x_k - T(x_k))]^2 \leq \lim_{k\to+\infty}[(T(x_k) - T^2(x_k))^t(x_k - T(x_k))]^2$. This result implies that $\lim_{k\to+\infty} \|x_k - T(x_k)\|^4 \leq \lim_{k\to+\infty} A^2\|x_k - T(x_k)\|^4$, but since $A < 1$, this is a contradiction, unless $\lim_{k\to+\infty} \|x_k - T(x_k)\|_G = 0$. Q.E.D.

Theorem 8 *: If $T = x - Mx + c$ is an affine map, with M and M^2 positive definite matrices, then the sequence $\{x_k\}$ that scheme V induces converges to the solution.*

Proof: See [23]. This result also follows from Theorem 1. Lemma 12 implies assumption A1. Moreover, since $a_k = \frac{d_k^t d_k}{d_k^t M d_k}$, assumptions A1 and A2 hold when M^2 and M are positive definite matrices.

To show that assumptions A1 and A2 are valid, we select $\bar{G} = \frac{M+M^t}{2}$. Then $A_k(x^*)$ becomes

$$A_k(x^*) = \frac{(x_k - x^*)^t M^2(x_k - x^*) + (x_k - x^*)^t M^t M(x_k - x^*)}{(x_k - x^*)^t M^t(\frac{M+M^t}{2})M(x_k - x^*)} - a_k =$$

(replacing $a_k = \frac{d_k^t d_k}{d_k^t M d_k}$),

$$\frac{(x_k - x^*)^t M^2(x_k - x^*) + d_k^t d_k}{d_k^t M d_k} - a_k = \frac{(x_k - x^*)^t M^2(x_k - x^*)}{d_k^t M d_k}.$$

Therefore, whenever M^2 is a positive definite matrix,

$$A_k(x^*) = \frac{(x_k - x^*)^t M^2(x_k - x^*)}{d_k^t M d_k} \geq$$

(for $x_k \neq x^*$),

$$\frac{\lambda_{min}(\frac{M^2+(M^2)^t}{2})}{\lambda_{max}(M^t G M)} = c > 0.$$

Q.E.D.

The following proposition characterizes the rate of convergence of scheme V.

Proposition 11 *: If $T = I - M$ is an affine mapping and M and M^2 are positive definite matrices and $\bar{M} = \frac{M+M^t}{2}$, then the sequence induced by scheme V contracts to a solution through the estimate,*

$$\|x_{k+1} - x^*\|_{\bar{M}}^2 \leq [1 - \frac{\lambda_{min}((\bar{M})^{-1}\bar{M^2})}{\lambda_{max}(\bar{M})}]\|x_k - x^*\|_{\bar{M}}^2. \tag{22}$$

Proof: See [23].

Example:

Let $K = R^n$ and $T(x) = [x_2, -x_1]$. Then $x^* = (0,0)$ is the solution of the fixed point problem $FP(T, R^n)$. The steepest descent algorithm starting from the point $x^0 = (1,1)$ generates the iterates $x^1 = (1,-1)$, $x^2 = (-1,-1)$, $x^3 = (-1,1)$, $x^4 = x^0 = (1,1)$ and, therefore, the algorithm cycles.

Remark:

In this example $M = I - T = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ is positive definite, but $M^2 = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$ is positive semidefinite. This example illustrates that the choice of potential we considered in scheme V could produce iterates that cycle unless the matrices M and M^2 are positive definite.

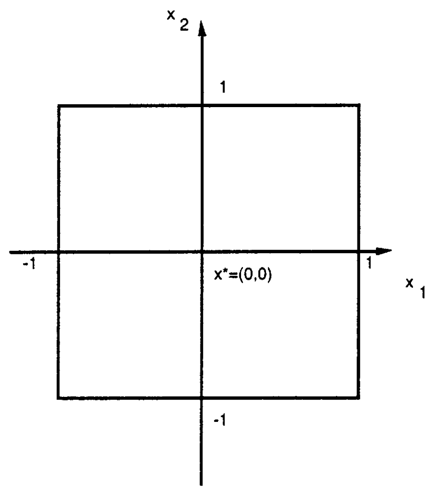

Figure 2: Cycling Example for Scheme V

Scheme VI: To remedy this cycling behavior, we modify the potential function in scheme V by introducing a penalty term as follows,

$$g_6(x_k(a)) = [(x_k(a) - T(x_k(a))^t G(x_k - T(x_k))]^2 -$$

$$\beta[(x_k(a) - x_k(0))^t G(x_k - T(x_k))].[(x_k(a) - x_k(1))^t G(x_k - T(x_k))].$$

We choose $S = R^+$.

We first make the following observation.

Lemma 13 *: If T is a nonlinear, nonexpansive map relative to the $||.||_G$ norm, then a choice of $\beta > 4$ implies that if $a_k \geq 1$, then x_k is a fixed point solution.*

Proof: The remark at the end of Section 2 implies that potential g_6 satisfies (i)-(v) in Subsection 2.2. Then this lemma follows from part *b* of Proposition 6. Q.E.D.

Remark:

Consider an *affine* map $T(x) = x - Mx + c$, and let $\bar{a}_k = \frac{\beta\|d_k\|_G^4 + 2\|d_k\|_G^2(d_k^t M^t G d_k)}{2[\beta\|d_k\|^4 + (d_k^t M^t G d_k)^2]}$. Then the step length solution a_k to the problem

$$min_{a\in[0,1]} g_6(x_k(a)),$$

is

$$a_k = min\{1, \bar{a}_k\}. \tag{23}$$

We observe the following,

(1) If $d_k \neq 0$, then $a_k \geq 1/2$.

(2) If $\bar{a}_k \geq 1$, then a choice of $\beta > 1$ implies that

$$0 \leq (d_k^t M^t G d_k)^2 + (\beta - 1)\|d_k\|_G^4 + (\|d_k\|_G^2 - d_k^t M^t G d_k)^2 \leq 0.$$

This result implies that $d_k = 0$ and, therefore, that x_k is a fixed point solution.

Lemma 14 *: Let T be a nonexpansive map relative to the G norm and suppose $\beta > 4$ in the general nonlinear case (or $\beta > 1$ in the affine case). If either a_k converges to zero or to one, then every limit point of the sequence $\{x_k\}$ is a fixed point solution.*

Proof: If a_k converges to zero, then

$$\lim_{k\to+\infty} g(x_k(0)) = \lim_{k\to+\infty} \|d_k\|_G^4 \leq \lim_{k\to+\infty} g(x_k(a)) \leq [1 - \beta a(1-a)] \lim_{k\to+\infty} \|d_k\|_G^4. \tag{24}$$

If $\lim_{k\to+\infty} d_k \neq 0$, then for all $a \in (0,1)$, $\lim_{k\to+\infty} g(x_k(0)) > [1 - \beta a(1-a)]\|d_k\|_G^4$, contradicting (24). We conclude that $\lim_{k\to+\infty} d_k = 0$ and therefore, every limit point of x_k is a fixed point solution. Moreover, if a_k converges to one, then if T is an affine map, if we choose $\beta > 1$,

$$0 \leq \lim_{k\to+\infty} (d_k^t M^t G d_k)^2 + (\beta - 1)\lim_{k\to+\infty} \|d_k\|_G^4 + \lim_{k\to+\infty} (\|d_k\|_G^2 - d_k^t M^t G d_k)^2 = 0.$$

This result implies that $\|d_k\|_G$ converges to zero. Therefore, Proposition 4 implies that every limit point of the sequence $\{x_k\}$ is a fixed point solution.

If T is a nonlinear, nonexpansive map then if a_k converges to one, then property *c.* of Proposition 6 implies that for a choice of $\beta > 4$, every limit point of the sequence $\{x_k\}$ is a fixed point solution. Q.E.D.

Theorem 9 *: If T is a nonexpansive map relative to the $\|.\|_G$ norm and $\beta > 4$ (or $\beta > 1$ in the affine case), then the sequence that scheme VI generates converges to a fixed point solution.*

Proof: Lemma 14 together with the nonexpansiveness of T imply that assumptions A1 and A2 are valid. The proof follows from Theorem 1. Q.E.D.

Remarks:

1) Observe that scheme VI extends the steepest descent method to include nonexpansive maps and does not require any positive definiteness condition on the square of the Jacobian matrix. Moreover, scheme VI provides a global convergence result even when the map T is nonlinear.

2) In this section we have considered special cases of the general scheme we introduced in Theorem 1. Several other schemes we have not presented in this section are also a special case of our general scheme. In particular, consider a variational inequality problem. Then Proposition 1 states that variational inequality $VI(f,K)$ is equivalent to a fixed point problem $FP(T,K)$. The operator T might be a projection operator $T = Pr_K^G(I - \rho G^{-1}f)$ or perhaps be defined so that $T(x) = y$, and y is a solution of a simpler variational inequality or minimization subproblem (see for example [12], [20], [44], [56]).

Fukushima [20] considered the variational inequality problem $VI(f,K)$. He considered the projection operator $T(x) = Pr_K^G(x - \rho G^{-1}f(x)) = argmin_{y\in K}[f(x)^t(y-x)+\frac{1}{2}\|y-x\|_G^2]$. The fixed point solutions corresponding to this map T are in fact the solutions of variational inequality $VI(f,K)$ (see Proposition 1). Using the potential

$$g(x) = -f(x)^t(T(x)-x) - \frac{1}{2}\|T(x)-x\|_G^2, \tag{25}$$

Fukushima established a scheme that, when f is strongly monotone, computes a variational inequality solution. Namely, each iteration of the scheme computes a point $x_{k+1} = x_k(a_k) = x_k + a_k(T(x_k) - x_k)$, with $a_k = argmin_{a\in[0,1]}g(x_k(a))$. We observe that this scheme is in fact a special case of our general scheme for the choice of potential g that we described in expression (25). Observe that when $f(x) = x - T(x)$, the potential in (25) is the same as the potential we considered in scheme II.

Taji, Fukushima and Ibaraki [52] established an alternative scheme for strongly monotone variational inequality problems, using an operator T that generates points through a Newton procedure. Zhu and Marcotte [58] modified Fukushima's scheme [20] to also include monotone problems. Wu, Florian and Marcotte [56] have generalized Fukushima's scheme. They considered as T the operator that maps a point x in the set $G(x) = argmin_{y\in K}[f(x)^t(y-x) + \frac{1}{\rho}\phi(x,y)]$. The function

$\phi : K \times K \rightarrow R$ has the following properties:

(a) ϕ is continuously differentiable,

(b) ϕ is nonnegative,

(c) ϕ is uniformly strongly convex with respect to y,

(d) $\phi(x, y) = 0$ is equivalent to $x = y$,

(e) $\nabla_x \phi(x, y)$ is uniformly Lipschitz continuous on K with respect to x.

Observe that when $\phi(x, y) = \|x - y\|_G^2$, then $T(x)$ becomes the projection operator as in Fukushima [20]. The fixed point solutions corresponding to this map T are also solutions of the variational inequality problem $VI(f, K)$ (see [56] for more details).

Furthermore, Wu, Florian and Marcotte [56] considered the potential,

$$g(x) = -f(x)^t(T(x) - x) - \frac{1}{\rho}\phi(T(x), x). \tag{26}$$

Notice as in the previous cases, this scheme becomes a special case of our general scheme for the choice of potential g given in (26).

The convergence of these schemes follows from Theorem 1, where we established the convergence of the general scheme. Observe that as we have established in Lemma 2 when F is strongly monotone (which the developers of these schemes impose) assumptions A1 and A2 hold for schemes which "repel" a_k from zero unless at a solution. Furthermore, observe that indeed this assumption follows for these schemes (like Lemma 7) using the descent property of the potential functions involved in each of the previous schemes. For a proof of the latter, see [20] and [56].

Finally, for all these schemes (see also Section 5), we can apply an Armijo-type procedure to compute inexact solutions to the line search procedures. This computation is easy to perform.

4 On the Rate of Convergence

In this section, we examine the following question: when does the general scheme we introduced in this paper and its special cases exhibit a better rate of convergence than the classic iteration (2)? To illustrate this possibility, we first establish a "best" step size under which the adaptive averaging scheme will achieve a better convergence rate.

Proposition 12 *: Let*

$$a_k^* = \frac{(x_k - T(x_k))^t \bar{G}(x_k - x^*)}{\|x_k - T(x_k)\|_{\bar{G}}^2} = \frac{(F(x_k) - F(x^*))^t \bar{G}(x_k - x^*)}{\|F(x_k) - F(x^*)\|_{\bar{G}}^2}.$$

If $a_k^* \neq 1$, then for some step size, namely a_k^*, the general iteration

$$x_{k+1} = x_k(a_k^*) = x_k + a_k^*(T(x_k) - x_k),$$

has a rate of convergence at least as good as the classic iteration (2). Moreover, if $|a_k^* - 1| \geq c > 0$, then it has a better rate of convergence.

Proof: First observe that

$$\|x_k(a) - x^*\|_{\bar{G}}^2 = \|x_k - x^*\|_{\bar{G}}^2 + a^2\|x_k - T(x_k)\|_{\bar{G}}^2 - 2a(x_k - T(x_k))^t\bar{G}(x_k - x^*). \tag{27}$$

If $x_k \neq T(x_k)$, then $\|x_k(a) - x^*\|_{\bar{G}}^2$ is a strictly convex function of a. Moreover,

$$a_k^* = argmin_a\|x_k(a) - x^*\|_{\bar{G}}^2 = \frac{(x_k - T(x_k))^t\bar{G}(x_k - x^*)}{\|x_k - T(x_k)\|_{\bar{G}}^2}.$$

Therefore, if $a_k^* \neq 1$, then

$$\|x_k(a_k^*) - x^*\|_{\bar{G}}^2 \leq \|T(x_k) - T(x^*)\|_{\bar{G}}^2 \leq A\|x_k - x^*\|_{\bar{G}}^2.$$

Moreover, expression (27) implies that

$$\|x_k(a_k^*) - x^*\|_{\bar{G}}^2 = \|T(x_k) - T(x^*)\|_{\bar{G}}^2 - (a_k^* - 1)^2\|x_k - T(x_k)\|_{\bar{G}}^2.$$

Proposition 3 implies that

$$\|x_k(a_k^*) - x^*\|_{\bar{G}}^2 \leq (A - (a_k^* - 1)^2(1 - \sqrt{A})^2)\|x_k - x^*\|_{\bar{G}}^2.$$

Therefore, if $|a_k^* - 1| \geq c > 0$, then

$$\|x_k(a_k^*) - x^*\|_{\bar{G}}^2 \leq (A - c^2(1 - \sqrt{A})^2)\|x_k - x^*\|_{\bar{G}}^2,$$

implying that the sequence $\{x_k(a^*)\}$ has a better rate of convergence than the sequence $\{T(x_k)\}$. Q.E.D.

Remark:

Which maps T give rise to step sizes $a_k^* \neq 1$, and which to step sizes $|a_k^* - 1| \geq c$? We next examine this question.

1. Let T be a nonexpansive map that is tight, that is,

$$\|T(x) - T(x^*)\|_{\bar{G}}^2 \approx \|x - x^*\|_{\bar{G}}^2.$$

Then

$$a_k^* = \frac{(x_k - T(x_k))^t \bar{G}(x_k - x^*)}{\|x_k - T(x_k)\|_{\bar{G}}^2} \approx \frac{1}{2}$$

and so $a_k^* - 1 \approx -\frac{1}{2}$. Observe that in this case, we can achieve the "best" rate of convergence by moving half way at each step.

2. Let T be a firmly nonexpansive map that is not tight, that is,

$$\|T(x) - T(x^*)\|_{\bar{G}}^2 < \|x - x^*\|_{\bar{G}}^2 - \|x - T(x) - x^* + T(x^*)\|_{\bar{G}}^2.$$

Setting $A = 1$, $B = -1$ and $y = x^*$ and $B = -1$ in Lemma 1 stated as a strict inequality implies that this condition is equivalent to

$$(x - T(x) - x^* + T(x^*))^t \bar{G}(x - x^*) > \|x - T(x)\|_{\bar{G}}^2.$$

Consequently, $a_k^* = \frac{(x_k - T(x_k))^t \bar{G}(x_k - x^*)}{\|x_k - T(x_k)\|_{\bar{G}}^2} > 1$.

3. Let T be a firmly contractive map. Then setting $y = x^*$, and $B = -1$ in Lemma 1 implies that

$$(x - T(x) - x^* + T(x^*))^t \bar{G}(x - x^*) \geq \frac{1 - A}{2}\|x - x^*\|_{\bar{G}}^2 + \|x - T(x)\|_{\bar{G}}^2.$$

Then Proposition 3 implies that $a_k^* - 1 \geq \frac{1 - \sqrt{A}}{2(1 + \sqrt{A})}$.

4. Finally, if T is contractive but with a constant A "close" to 1, that is, if for some constant $\bar{A} \in (0, 1)$,

$$A\|x - x^*\|_{\bar{G}}^2 \geq \|T(x) - T(x^*)\|_{\bar{G}}^2 \geq \|x - x^*\|_{\bar{G}}^2 - \bar{A}\|x - T(x)\|_{\bar{G}}^2,$$

then $1 - a_k^* \geq \frac{1 - \bar{A}}{2} > 0$.

The previous analysis shows that for certain types of maps and for some step sizes, adaptive averaging provides a better rate of convergence than the classical iterative scheme $x_{k+1} = T(x_k)$. We have shown that to achieve the improved convergence rate, we need to choose step size as a_k^*. Since a_k^* involves a fixed point solution x^*, its computation is not possible. Accordingly, we extend our choice of step sizes to establish an allowable range of step sizes for which adaptive averaging schemes have a better rate of convergence. Moreover, we show that the schemes we studied in the previous section have step sizes within this range.

Proposition 13 : *For fixed point problems with maps T satisfying the condition*

$$|a_k^* - 1| = \left|\frac{(x - T(x))^t \bar{G}(x - x^*)}{\|x - T(x)\|_{\bar{G}}^2} - 1\right| \geq c > 0, \quad \text{for all } x \in K \tag{28}$$

adaptive averaging schemes with choices of step size a_k lying with the range

$$a_k^* - |a_k^* - 1| + d \leq a_k \leq a_k^* + |a_k^* - 1| - d, \tag{29}$$

with $0 < d < c$, have a better rate of convergence than the classic iteration (2).

Proof: Using expression (27) we see that

$$\|x_k(a) - x^*\|_G^2 = \|T(x_k) - T(x^*)\|_G^2 + (a^2 - 2aa_k^* + 2a_k^* - 1)\|x_k - T(x_k)\|_G^2.$$

Therefore, the binomial $a^2 - 2aa_k^* + 2a_k^* - 1 < -d^2$ for choices of step sizes a satisfying the condition

$$a_k^* - |a_k^* - 1| + d \leq a \leq a_k^* + |a_k^* - 1| - d,$$

with $0 < d < c$. Consequently, we need $|a_k^* - 1| \geq c > 0$, (that is, condition (28)). Functions of the type we discussed in the previous remark satisfy this inequality. Then a choice of step sizes a_k lying within the range of (29) provide a rate of convergence

$$\|x_k(a_k) - x^*\|_G^2 \leq \|T(x_k) - T(x^*)\|_G^2 - d^2\|x_k - T(x_k)\|_G^2$$

(Using the fact that T is contractive and Proposition 3, we obtain)

$$\|x_k(a_k) - x^*\|_G^2 \leq [A - d^2(1 - \sqrt{A})^2]\|x_k - x^*\|_G^2. \tag{30}$$

Q.E.D.

Corollary 7 *: For fixed point problems with maps T satisfying expression (28), a choice of step sizes within (29) guarantees a better rate of convergence than the classic iteration (2). Furthermore,*

$$\|x_{k+1} - x^*\|_G^2 \leq \|T(x_k) - T(x^*)\|_G^2 - (a_k - 1)(A_k(x^*) - 1)\|x_k - T(x_k)\|_G^2 \leq \tag{31}$$

$$[A - (a_k - 1)(A_k(x^*) - 1)(1 - \sqrt{A})^2]\|x_k - x^*\|_G^2.$$

Proof: Expression (31) follows similarly to our development of (7). Moreover, observe that $A_k(x^*) = 2a_k^* - a_k$. Therefore, to obtain a better rate of convergence than the classical iteration (2), we need $(a_k - 1)(A_k(x^*) - 1)$ bounded away from 0. Since, $(a_k - 1)(A_k(x^*) - 1) = a_k^2 - 2a_k.a_k^* + 2a_k^* - 1$, the conclusion follows as in Proposition 13. Q.E.D.

Remarks:

1) Condition (29) is equivalent to assuming that $(a_k - 1)(A_k(x^*) - 1) \geq d^2 > 0$. Moreover, it is equivalent to assuming that

$$A_k(x^*) \in [min\{1, 2a_k^* - 1\} + d, max\{1, 2a_k^* - 1\} - d]. \tag{32}$$

Observe that when (29) (or (32)) hold, then assumptions A1 and A2 also follow.

2) In the following discussion, we show that the step sizes that we used in the schemes of Section 3 satisfy (29). Therefore, the schemes we studied in Section 3 exhibit better rates of convergence than the classical iteration (2). In discussing these methods, we let a_k^j and $A_k^j(x^*)$ with $j = I, ..., VI$ denote the step size and quantity $A_k^j(x^*)$ for scheme j at the kth iteration.

In particular, consider the affine problem with $I - T = M$. We need to keep in mind that $T(x_k) - x_k = d_k$ and $a_k^* = \frac{d_k^t \bar{G}(x^* - x_k)}{\|d_k\|_{\bar{G}}^2}$.

• Consider scheme I. If T is firmly contractive, then relation (29) follows. If we set $\bar{G} = M^t M + \beta I$, then $a_k^* = \frac{d_k^t M d_k + \beta(x_k - x^*)^t M(x_k - x^*)}{\beta\|d_k\|^2 + \|Md_k\|^2}$. Furthermore, for $G = I$, $a_k^I = \frac{d_k^t M d_k}{\beta\|d_k\|^2 + \|Md_k\|^2}$.

Observe that for firmly contractive maps T, a choice of $\beta < \frac{1-A}{2}$ implies, using Proposition 3, that

$$a_k^I - 1 \geq \frac{1 - A - 2\beta}{2} \frac{\|d_k\|^2}{\|Md_k\|^2 + \beta\|d_k\|^2} \geq \frac{1 - A - 2\beta}{2(\beta + (1 + \sqrt{A})^2)} = d > 0.$$

Therefore,

$$(a_k^I - 1)(A_k^I(x^*) - 1) = (a_k^I - 1)(2a_k^* - a_k^I - 1) \geq (a_k^I - 1)^2 \geq d^2,$$

with $d = \frac{1-A-2\beta}{2(\beta+(1+\sqrt{A})^2)}$, that is, relation (29).

• Consider scheme II. If T is contractive, then relation (29) follows. Observe that a choice of $\bar{G} = M^t M$ in a_k^* and $G = I$ in a_k^{II} implies that $a_k^{II} = a_k^*$ which satisfies relation (29).

• Consider scheme III. If T is contractive, then relation (29) follows. If we set $\bar{G} = M^t M - \beta I$ then $a_k^* = \frac{d_k^t M d_k - \beta(x_k - x^*)^t M(x_k - x^*)}{\|Md_k\|^2 - \beta\|d_k\|^2}$. Furthermore, for $G = I$,
$a_k^{III} = \frac{d_k^t M d_k}{\|Md_k\|^2 - \beta\|d_k\|^2}$, if $\|Md_k\|^2 > \beta\|d_k\|^2$ and $\frac{d_k^t M d_k}{\|Md_k\|^2 - \beta\|d_k\|^2} < c_1$. Otherwise $a_k^{III} = c_1$.
Observe that $1 - a_k^{III} \geq 1 - c_1 = d > 0$. Then

$$1 - A_k^{III}(x^*) = \frac{\|Md_k\|^2 - \beta\|d_k\|^2 - d_k^t M d_k + 2\beta(x_k - x^*)^t M(x_k - x^*)}{\|Md_k\|^2 - \beta\|d_k\|^2} \geq 1 - c_1 = d.$$

Therefore, $(1 - a_k^{III})(1 - A_k^{III}(x^*)) \geq d^2$, that is, relation (29).

• Consider scheme IV. If T is firmly contractive, then relation (29) follows. If we set $G = I$, then $a_k^{IV} = \frac{\beta\|d_k\|^2 + 2d_k^t M d_k}{2\beta\|d_k\|^2 + 2\|Md_k\|^2}$. If T is a firmly contractive map, Lemma 1 with $B = -1$ and Proposition 3 imply that

$$a_k^{IV} - 1 \geq (1 - A - \beta)\frac{\|d_k\|^2}{2(\beta\|d_k\|^2 + \|Md_k\|^2)} \geq (1 - A - \beta)\frac{1}{2(\beta + (1 + \sqrt{A})^2)}.$$

Therefore, a choice of $\beta < 1 - A$ guarantees that $a_k^{IV} - 1 \geq d = (1 - A - \beta)\frac{1}{2(\beta+(1+\sqrt{A})^2)} > 0$.

Moreover, observe that for a choice of $\bar{G} = M^t M$, $a_k^* = \frac{d_k^t M d_k}{\|Md_k\|^2}$. If T is firmly contractive, then

$$A_k^{IV}(x^*) - 1 = 2a_k^* - a_k^{IV} - 1 = \frac{2d_k^t M d_k - \|Md_k\|^2}{\|Md_k\|^2} - a_k^{IV} \geq$$

$$\frac{d_k^t M d_k + \frac{1-A}{2}\|d_k\|^2}{\|Md_k\|^2} - a_k^{IV} \geq \frac{(\frac{1-A-\beta}{2})\|d_k\|^2}{\|Md_k\|^2} \geq d,$$

for all $\beta < 1 - A$. Therefore, a choice of $\beta < 1 - A$ guarantees that $a_k^{IV} - 1)(A_k^{IV}(x^*) - 1) \geq d^2$ with $d = (1 - A - \beta)\frac{1}{2(\beta+(1+\sqrt{A})^2)} > 0$, that is, relation (29).

- Consider scheme V, that is, the steepest descent method. If T is contractive and M is symmetric, then relation (29) follows. When $M = M^t$ then a choice of $\bar{G} = M$ and $G = I$ imply that $a_k^* = a_k^V$, which satisfies relation (29).
- Consider scheme VI. If T is contractive and M is symmetric, then relation (29) follows. Let us choose $\beta \geq \frac{(1+\sqrt{A})^4+1}{4c}$ with $c > 1$. Then for $G = I$,

$$a_k^{VI} = \bar{a}_k^{VI} = \frac{\beta\|d_k\|^4 + 2\|d_k\|^2(d_k^t M^t d_k)}{2[\beta\|d_k\|^4 + (d_k^t M^t d_k)^2]} \geq c > 1.$$

Therefore, $a_k^{VI} - 1 \geq c - 1 = d > 0$. Furthermore, if $\bar{G} = \frac{M+M^t}{2}$, then $A_k^{VI}(x^*) - 1 = 2a_k^* - a_k^{VI} - 1 \geq a_k^{VI} - 1 \geq c - 1 = d > 0$. This condition follows since

$$a_k^* = \frac{d_k^t d_k}{d_k^t M d_k} \geq \frac{\beta\|d_k\|^4 + 2\|d_k\|^2(d_k^t M^t d_k)}{2[\beta\|d_k\|^4 + (d_k^t M^t d_k)^2]} = a_k^{VI}.$$

This inequality is valid since $d_k^t M d_k \leq \|d_k\|.\|Md_k\| \leq (1 + \sqrt{A})\|d_k\|^2 \leq 2\|d_k\|^2$.

Finally, we need to consider how our previous results extend for the general nonlinear case. That is, when does the general scheme we introduced in Section 2 as well as the special cases we studied in Section 3 satisfy relation (29) for general nonlinear operators T? In this case, these schemes would then demonstrate a better rate of convergence than the classical iteration (2).

Observe that if for example $a_k - 1 \geq c - 1 = d > 0$, then in order to establish (29), we need to find for which choice of $\bar{G}$ and β and for what maps T, step sizes $a_k \leq a_k^*$?

5 Inexact Line searches

One natural question would arise when attempting to implement any of the methods we have examined in Section 3:

how easy is it to perform line searches ? The line searches might not be easy to perform in the general nonlinear case. For this reason, in this section we consider inexact line searches.

We determine step sizes at each step by applying an Armijo-type rule of the following form. For positive constants $D > 0$ and $0 < b < 1$, find the smallest integer l_k so that a step length $a_k = b^{l_k}$ satisfies the condition

$$g_i(x_k(a_k)) - g_i(x_k(0)) \leq -D.b^{l_k}\|d_k\|_G^2. \tag{33}$$

In this expression, g_i denotes the potential function we have used previously for scheme i. We next show that the Armijo-type inexact line search (33) works for the various choices of potentials we considered in Section 3, if we impose appropriate assumptions on the fixed point map T. Since the potentials we considered in Section 3 involve a term $\|x_k(a) - T(x_k(a)\|_G^2$ we first make some preliminary observations concerning this term that is, we first examine scheme II.

Theorem 10 : *Consider a fixed point problem $FP(T, K)$. Let T be a contractive map relative to the $\|.\|_G$ norm with a coercivity constant $A \in (0, 1)$. Then a choice of $D \geq 1 - \sqrt{A}$ in an Armijo-type search (33) applied to the potential $g_2(x(a)) = \|x(a) - T(x(a))\|_G^2$ generates a sequence that converges to a fixed point solution.*

Proof: We first need to show that the Armijo-type line search (33) has a solution. To see why this is true, we observe that the contractiveness of the map T and Corollary 4 imply that all $0 \leq a \leq 1$ satisfy the inequality

$$g_2(x_k(a)) - g_2(x_k) \leq -(1 - \sqrt{A})a\|d_k\|_G^2.$$

Therefore, if we choose $D \geq 1 - \sqrt{A}$, then all step lengths $0 \leq a \leq 1$ satisfy the Armijo-type rule (33).

If in the Armijo-type inequality (33) we choose a step length $a_k = b^{l_k} < 1$, so that l_k is the smallest integer satisfying (33), then $1 \geq a_k \geq b$. which implies that

$$\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \geq a_k^2\|x_k - T(x_k)\|^2. \tag{34}$$

The observation that $1 \geq a_k \geq b$ implies that

$$\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \geq (b)^2\|x_k - T(x_k)\|^2.$$

This result in turn implies that the sequence $\|x_k - T(x_k)\|^2$ converges to zero. Proposition 4 implies that the entire sequence $\{x_k\}$ converges to a fixed point solution. Q.E.D.

Theorem 11 *: Consider fixed point problems $FP(T, K)$ with nonexpansive maps T relative to the $\|.\|_G$ norm. If we apply Armijo-type (33) line searches on potential functions g_i, $i = 3, 4$, then the sequence x_k that these line searches generate converges to a fixed point solution.*

Proof: We first need to show that the Armijo-type (33) line searches have a solution. In Corollary 4 we have shown that for nonexpansive maps T,

$$\|x_k(a) - T(x_k(a)\|_G^2 - \|x_k - T(x_k)\|_G^2 \leq 0,$$

for all $0 \leq a \leq 1$. Applying this result to potential functions of the type

$$g_i(x_k(a)) = \|x_k(a) - T(x_k(a)\|_G^2 - \beta h_i(a)\|x_k - T(x_k)\|_G^2,$$

implies that

$$g_i(x_k(a)) - g_i(x_k) \leq -\beta h_i(a)\|x_k - T(x_k)\|_G^2, \tag{35}$$

for all $0 < a < 1$.

Therefore, we see that the step sizes a_k will satisfy Armijo-type line searches condition (33) as long as $\beta h_i(a_k) \geq D.a_k$.

To complete the proof, we next consider specific choices of $h_i(a)$ that correspond to the various choices of potentials g_i, for $i = 3, 4$ that we considered in Section 3.

1) Consider the potential $g_3(x_k(a)) = \|x_k(a) - T(x_k(a)\|_G^2 - \beta\|x_k(a) - x_k(0)\|_G^2$. As we have shown in Theorem 5 we consider only step sizes $a < c_1$. We need to select constants D and β so that $\frac{D}{\beta} < c_1 < 1$. In this case it is easy to see that since $h_3(a) = a^2$, all step sizes $c_1 \geq a \geq \frac{D}{\beta}$ satisfy the inequality $\beta h_i(a) \geq D.a$. As a result, these step sizes satisfy the inexact Armijo-type inequality (33) as well.

Choosing $a_k = b^{l_k}$, so that l_k is the smallest integer satisfying (33) implies that $a_k \geq b$. Therefore, $max(b, D/\beta) \leq a_k < c_1$. Convergence follows since $max(b, D/\beta) \leq a_k \leq c_1$ implies, as we argued in Theorem 1, that

$$\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \geq c.a_k^2\|x_k - T(x_k)\|^2 \geq c.max(b, D/\beta)^2\|x_k - T(x_k)\|^2, \tag{36}$$

with $c = \frac{1-c_1}{c_1}$. Expression (36) implies that the sequence $\{x_k\}$ converges to a fixed point solution.

2) Consider the potential $g_4(x_k(a)) = \|x_k(a) - T(x_k(a)\|_G^2 - \beta\|x_k(a) - x_k(0)\|.\|x_k(a) - x_k(1)\|$. Then $h_4(a) = a(1-a)$, implying that for a choice of $\beta > D$, all $0 \leq a \leq 1 - D/\beta$ satisfy the Armijo-type inequality (33).

Moreover, choosing $a_k = b^{l_k}$, with l_k the smallest integer satisfying (33) implies that $a_k \geq b$. Therefore, the Armijo step size a_k lies in the interval $[b, 1 - D/\beta]$. As we have shown in Theorem 7,

$$\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \geq a_k(1 - a_k)\|x_k - T(x_k)\|^2 \geq b(1 - D/\beta)\|x_k - T(x_k)\|^2. \tag{37}$$

Similar arguments to those used in the proof of Theorem 7, imply then that the sequence $\{x_k\}$ converges to a fixed point solution. Q.E.D.

6 Conclusions-Open Questions

In this paper we introduced adaptive averaging schemes for solving fixed point problems. They allowed us to show how to solve classes of fixed point problems whose maps expand in some way, that is, are weaker than nonexpansive. We considered a general scheme for determining step sizes dynamically by optimizing a potential function. We considered several choices of potential functions that optimized in some sense "how far" the current point lies from the image of the fixed point mapping of the current point. We established convergence rates for these choices of potential functions. Moreover, we studied when our general scheme has a better rate of convergence than the classic iteration (2). The line searches we proposed might be hard to perform exactly. For that reason, we also considered inexact line searches.

Several open questions that naturally follow from our analysis,

• *How does the behavior of the schemes we introduced compare for the various choices of potentials in practice?*

• *It might be preferable to consider averages of the type* $x_k = \frac{a^1 x_1 + a^2 T(x_1) + \ldots + a^k T(x_{k-1})}{a^1 + \ldots + a^k}$, *and optimize potentials involving all* a^i*'s, instead of moving along* $x_k(a) = x_k + a_k(T(x_k) - x_k)$ *and optimize potentials involving only* $a_k = \frac{a^k}{a^1 + \ldots + a^k}$.

• *Would these results still be valid if we impose conditions that are weaker than nonexpansiveness?*

• *Can we establish rates of convergence for other choices of potentials?*

References

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

[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 (1976), "On the Goldstein-Levitin-Polyak gradient projection method", *IEEE Transactions on Automatic Control*, 21, 174-184.

[4] D.P. Bertsekas (1982), "Projected Newton methods for optimization problems with simple constraints", *SIAM Journal on Control and Optimization*, 20, 221-246.

[5] D.P. Bertsekas (1996), *Nonlinear programming*, Athena Scientific, Belmont, MA.

[6] K.C. Border (1989), *Fixed point theorems with applications to economics*, Cambridge University Press.

[7] F.E. Browder (1967), Convergence of approximants to fixed points of nonexpansive nonlinear mappings in Banach spaces", *Arch. Rational Mech. Anal.*, 24, 82-90.

[8] F.E. Browder and W.V. Petryshn (1967), "Construction of fixed points of nonlinear mappings in a Hilbert space", *Journal of Analysis and Applications*, 20, 197-228.

[9] R.E. Bruck, Jr. (1973), "The iterative solution of the equation $y \in x + T(x)$ for a maximal monotone map T in Hilbert space", *Journal of Mathematical Analysis and Applications*, 48, 114-126.

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

[11] P.H. Calamai and J.J. More (1987), "Projected gradient methods for linearly constrained problems", *Mathematical Programming*, 39, 93-116.

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

[13] J.C. Dunn (1973), "On recursive averaging processes and Hilbert space extensions of the contraction mapping principle", *Journal of the Franklin Institute*, 295, 117-133.

[14] J.C. Dunn (1980), "Convergence rates for conditional gradient sequences generated by implicit step length rules", *SIAM Journal on Control and Optimization*, 18, 5, 473-487.

[15] J.C. Dunn (1980), "Newton's method and the Goldstein step length rue for constrained minimization problems", *SIAM Journal on Control and Optimization*, 18, 659-674.

[16] J.C. Dunn (1981), "Global and asymptotic convergence rate estimates for a class of projected gradient processes", *SIAM Journal on Control and Optimization*, 19, 3, 368-400.

[17] J.C. Dunn and S. Harshbarger (1978), "Conditional gradient algorithms with open loop step size rules", *Journal of Mathematical Analysis and Applications*, 62, 432-444.

[18] J. Eckstein (1989), *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, Ph.D. thesis, Laboratory for Information and Decision Systems, MIT, Cambridge, MA.

[19] 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.

[20] M. Fukushima (1992), "Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems", *Mathematical Programming*, 53, 1, 99-110.

[21] 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.

[22] B. Halpern (1967), "Fixed points of nonexpansive maps", *Bulletin of the American Mathematical Society*, 73, 957-961.

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

[24] P.T. Harker (1993), "Lectures on computation of equilibria with equation-based methods, *CORE Lecture Series*, Center for Operations Research and Econometrics, University of Louvain, Louvain la Neuve, Belgium.

[25] 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.

[26] T. Hoang (1991), "Computing fixed points by global optimization methods", in *Fixed point theory and applications*, (M.A. Thera and J-B Baillon Eds.), 231-244.

[27] S. Ishikawa (1976), "Fixed points and iteration of a nonexpansive mapping in a banach space", *Proceedings of the American Mathematical Society*, 59, 1, 65-71.

[28] T. Itoh, M. Fukushima and T. Ibaraki (1988), "An iterative method for variational inequalities with application to traffic equilibrium problems", *Journal of the Operations Research Society of Japan*, 31, 1, 82-103.

[29] S. Kaniel (1971), "Construction of a fixed point for contractions in banach space", *Israel J. Math.*, 9, 535-540.

[30] W.A. Kirk (1990), *Topics in metric fixed point theory*, Cambridge Studies in Advanced Mathematics, 28, Cambridge University Press, New York.

[31] W.A. Kirk (1991), "An application of a generalized Krasnoselskii-Ishikawa iteration process", in *Fixed point theory and applications*, (M.A. Thera and J-B Baillon Eds.), 261-268.

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

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

[34] A.N. Kolmogorov and S.V. Fomin (1970), *Introduction to real analysis*, Dover Publications, New York.

[35] I. Konnov (1994), "A combined relaxation method for variational inequalities with nonlinear constraints", preprint.

[36] T.L. Magnanti and G. Perakis (1995), "A unifying geometric solution framework and complexity analysis for variational inequalities", *Mathematical Programming*, 71, 327-351.

[37] T.L. Magnanti and G. Perakis (1997), "The orthogonality theorem and the strong-f-monotonicity condition for variational inequality algorithms", *SIAM Journal of Optimization*, 7, 1, 248-273.

[38] T.L. Magnanti and G. Perakis (1994), "Averaging schemes for variational inequalities and systems of equations", *Mathematics of Operations Research*, 22, 3, 568-587.

[39] W.R. Mann (1953), "On recursive averaging processes and Hilbert space extensions of a contraction mapping principle", *Proceedings of the American Mathematical Society*, 4, 506-510.

[40] P. Marcotte and J-P Dussault (1987), "A note on a globally convergent Newton method for solving monotone variational inequalities", *Operations Research Letters*, 6, 1, 35-42.

[41] P. Marcotte and J.H. Wu (1995), "On the convergence of projection methods: application to the decomposition of affine variational inequalities", *JOTA*, 85, 2, 347-362.

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

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

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

[45] J.S. Pang (1994), "Complementarity problems", *Handbook of Global Optimization*, eds. R. Horst and P. Pardalos, 271-329, Kluwer Academic Publishers.

[46] 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.

[47] G. Perakis (1993), *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.

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

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

[50] R.C. Sine (1983), "Fixed Points and Nonexpansive Mappings", *Contemporary Mathematics*, 18, American Mathematical Society.

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

[52] K. Taji, M. Fukushima and T. Ibaraki (1993), "A globally convergent Newton method for solving strongly monotone variational inequalities", *Mathematical Programming*, 58, 3, 369-383.

[53] M.J. Todd (1976), " The computation of fixed points and applications" *Lecture Notes in Economics and Mathematical Systems*, 124, Springer-Verlag.

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

[55] P. Tseng, D.P. Bertsekas and J.N. Tsitsiklis (1990), "Partially asynchronous, parallel algorithms for network flow problems and other problems", *SIAM Journal on Control and Optimization*, 28, 3, 678-710.

[56] J.H. Wu, M. Florian and P. Marcotte (1993), "A general descent framework for the monotone variational inequality problem", *Mathematical Programming*, 61, 3, 281-300.

[57] D.L. Zhu and P. Marcotte (1996), "Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities", *SIAM Journal on Optimization*, 3, 6, 714-726.

[58] D.L. Zhu and P. Marcotte (1993), "Modified descent methods for solving the monotone variational inequality problem", *Operations Research Letters*, 14, 2, 111-120.

TABLE I

Potential	Convergence Conditions	Rate-$P(x_k, T(x_k))$	Rate-$\|x_k - x^*\|_G^2$
$P(x_k(a), T(x_k(a)) - \beta h(a) P(x_k, T(x_k)),\ S \subseteq R^+$	(see Proposition 5)	$1 - \beta(C - h(a_k))$	$A - d^2(1-\sqrt{A}),\ d < \|a_k^* - 1\|$
	if T firm contr., $d < \frac{1-\sqrt{A}}{2(1+\sqrt{A})}$		$A - d^2(1-\sqrt{A})^2$
(I) $\|x_k(a) - T(x_k(a))\|_G^2 + \beta\|x_k(a) - x_k(0)\|_G^2,\ S = R^+$	T contract., $\beta < 1 - A$, or	$A - \beta(a_k^2 - 1)$	"
	$T = I - M$, nonexp.	"	"
(II) $\|x_k(a) - T(x_k(a))\|_G^2,\ S = R^+$	T contract., or $T = I - M$, nonexp.	A	"
(III) $\|x_k(a) - T(x_k(a))\|_G^2 - \beta\|x_k(a) - x_k(0)\|_G^2,\ S = [0, c_1],\ c_1 \leq 1$	T nonexp., $c_1 < 1$.	$1 - \beta(c_1^2 - a_k^2)$	"
(IV) $\|x_k(a) - T(x_k(a))\|_G^2 - \beta\|x_k(a) - x_k(0)\|_G \cdot \|x_k(a) - x_k(1)\|_G,\ S = R^+$	T nonexp.	$1 - \beta(a(k) - \frac{1}{2})^2$	"
(V) $[(x_k(a) - T(x_k(a)))^t(x_k - T(x_k))]^2,\ S = R^+$	$T = I - M$, M and M^2 pos. def.		$1 - \frac{\lambda_{min}((\bar{M})^{-1}\bar{M}^2)}{\lambda_{max}(M)}$
(VI) $[(x_k(a) - T(x_k(a)))^t(x_k - T(x_k))]^2$			
- $\beta[(x_k(a) - x_k(0))^t(x_k - T(x_k))].[(x_k(a) - x_k(1))^t(x_k - T(x_k))],\ S = R^+$	T nonexp., $\beta > 4$		"