Math. Program., Ser. A 101: 435–461 (2004)

Digital Object Identifier (DOI) 10.1007/s10107-003-0476-5

Thomas L. Magnanti · Georgia Perakis

Solving variational inequality and fixed point problems by line searches and potential optimization

Received: January 5, 1998 / Accepted: October 9, 2003
Published online: 20 August 2004 –

Abstract. We introduce a general adaptive line search framework for solving fixed point and variational inequality problems. Our goals are to develop iterative 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 introduce and study joint compatibility conditions imposed upon the underlying map and the iterative step sizes at each iteration and 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

1. Introduction

Fixed point and variational inequality problems define two closely related and broad classes of problems that arise in the context of optimization as well as in fields as diverse as economics, game theory, transportation science, and regional science. Their widespread applicability motivates the need to develop and study efficient solution algorithms.

The variational inequality problem that we consider

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

is 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 literature for solving variational inequalities is vast. Review papers by Harker and Pang [25], Pang [43] and Florian and Hearn [19], and books by Harker [24], Nagurney [39] and more recently by Facchinei and Pang [18] summarize and categorize many algorithms for the problem as well as the application of these methods to special settings such as the traffic equilibrium problem. Often algorithms for solving variational inequalities establish an algorithmic map $T : K \subseteq R^n \to K$ and a fixed point problem,

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

T.L. Magnanti: School of Engineering and Sloan School of Management MIT, Cambridge, MA 02139

G. Perakis: Sloan School of Management and Operations Research Center, MIT, Cambridge, MA 02139

Preparation of this paper was supported, in part, from the National Science Foundation NSF Grant 9634736-DMI, as well as the Singapore-MIT Alliance

whose solution solves the original problem. Examples of algorithmic maps include a projection operator $T = Pr_K(I - \rho f)$ for some positive constant ρ or, more generally, a map $T(x)$ determined by solving a simpler variational inequality or minimization subproblem (see, for example, [11] and [42]). Moreover, often algorithms for solving variational inequality problems in various applied settings are related to the solution of fixed point problems (see, for example, [10], [24], [39]).

A standard method for solving fixed point problems for contractive maps T is to apply function iteration, $x_{k+1} = T(x_k)$. The classical Banach fixed point theorem shows that for contractive maps this method converges from any starting point to the unique fixed point of T. When the map is not contractive (e.g, is nonexpansive), function iteration need not converge and, indeed, the map T need not have a fixed point (or it might have several). For example, when $T(y, z) = b(z, -y)$, for some constant b, then the sequence that function iteration induces (i) does not converge to the solution $x^* = 0$ when $b = 1$, but rather induces a 90^o degree rotation about this point, and (ii) when $b > 1$, moves away from the solution. Is there a way to remedy this cycling and nonconvergent behavior?

In the case of cycling behavior, researchers (see [1], [8], [12], [27], [35], [36], [40], [44]) have established convergence of recursive averaging (line search) schemes of the type

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

assuming that $0 < a_k < 1$ and $\sum_{k=1}^{\infty} a_k(1-a_k) = +\infty$. We will refer to this sequence of step sizes as a Dunn sequence. 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 [35]).

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 function iteration.

Motivated by these observations, in this paper we introduce and study a recursive line search framework for solving fixed point and variational inequality problems. Our goals are to (i) design methods permitting a larger range of step sizes and/or better rates of convergence than prior methods (for example, function iteration 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:

1. Introduce, in Section 2, a general adaptive line search framework that relies on step sizes that are "compatible" with a given problem map T. The convergence results from this section highlight, and quantify, tradeoffs inherent in imposing properties (e,g., contractiveness assumptions) on the underlying fixed point map and restrictions on the iterative step sizes. The convergence results include as special cases, classical

function iteration for contractive maps and various averaging methods (Dunn's averaging results ([12], [35])) for nonexpansive maps.
2. Introduce and study, in Section 3, several line search procedures that dynamically optimize potential functions and, in doing so, "intelligently" choose step sizes. These methods are special, implementable, cases of the general framework developed in Section 2. For example, to determine step sizes for nonexpansive maps, we generalize the condition mentioned previously for ensuring convergence of certain (Dunn) averaging methods: unless the current iterate lies close to a fixed point solution, the next iterate should not lie close to either the current iterate or its image under the fixed point map T.

As a side benefit of our analysis, we are able to show that a modification of the classical steepest descent method for solving fixed point problems (or, equivalently, unconstrained asymmetric variational inequality problems) 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. The method is globally convergent even in the general nonlinear case.

In Section 4, we show that under appropriate conditions, the general framework and the potential methods we examine have better rates of convergence than function iteration and yield a "close to optimal" rate of convergence. In Section 5, as an aid to implementation, we consider inexact line searches. Finally, in Section 6, we state some open questions.

1.1. Preliminaries

We will impose several conditions on the underlying fixed point map T. In the statement of these conditions and elsewhere in this paper, G or $\bar{G}$ denotes a positive definite matrices.

Definition 1. *A map $T : K \to K$ is* **Lipschitz continuous** *relative to the $\|.\|_G$ norm on the set K if for some* **Lipschitz** *constant $A > 0$,*

$$\|T(x) - T(y)\|_G^2 \leq A\|x - y\|_G^2 \quad \forall x, y \in K. \tag{3}$$

If x^* is a fixed point solution, then the map T is **Lipschitz continuous** *(contractive or nonexpansive respectively)* **around the solution** *x^* if*

$$\|T(x) - T(x^*)\|_G^2 \leq A\|x - x^*\|_G^2 \quad \forall x \in K.$$

Special cases

When $0 < A < 1$, the map T is a **contraction on** *K relative to the $\|.\|_G$ norm, and when $A = 1$, the map T is* **nonexpansive on** *K, relative to the $\|.\|_G$ norm.*

When a map T is a contraction, we will sometimes refer to the constant A as the **contraction** *constant.*

The map $T(y, z) = b(z, -y)$ that we introduced previously is nonexpansive when $b = 1$ and contractive when $0 < b < 1$.

The following bounds will be useful at later points of the paper.

Proposition 1. *If T is a Lipschitz continuous relative to the $\|.\|_G$ norm with a Lipschitz constant $A > 0$, x^* is a fixed point of T, and $x \neq x^*$, then the following inequalities are valid:*

$$(1 + \sqrt{A})\|x - x^*\|_G \geq \|x - T(x)\|_G \geq (1 - \sqrt{A})\|x - x^*\|_G. \qquad (4)$$

Proof. Lipschitz continuity implies that

$$\begin{aligned} ||x - x^*||_G^2 + ||x - T(x)||_G^2 - 2(x - T(x))^t G(x - x^*) &= ||T(x) - x^*||_G^2 \\ &= ||T(x) - T(x^*)||_G^2 \\ &\leq A||x - x^*||_G^2. \end{aligned}$$

From this expression, the inequality $|(x - T(x))^t G(x - x^*)| \leq \|x - T(x)\|_G \|x - x^*\|_G$ implies that

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

If $x = x^*$, then the inequalities (4) are trivially valid. Otherwise, dividing both sides of the previous inequality by $\|x - x^*\|_G^2$ and setting $w = \frac{\|x-T(x)\|_G}{\|x-x^*\|_G}$, we obtain the valid inequality $\frac{1}{2}w^2 - w + \frac{1-A}{2} \leq 0$. This inequality holds only for values of w lying between the two roots $w_1 = 1 + \sqrt{A}$ and $w_2 = 1 - \sqrt{A}$ of the binomial, which implies the inequality (4). □

As a final preliminary, we introduce a definition and a simple proposition that relate limit points of a sequence to fixed point solutions.

Definition 2. *The sequence $\{x_k\}$ is asymptotically regular (with respect to the map T) if $\lim_{k\to\infty} \|x_k - T(x_k)\|_G = 0$, for some positive definite matrix G.*

Proposition 2. *(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 T is a continuous map and the sequence $\{x_k\}$ is asymptotically regular, then every limit point, if any, of this sequence is also a fixed point solution.

Proof. Property (i) is a direct consequence of continuity since if the sequence converges to some fixed point x^*, then the continuity of T implies that $\lim_{k\to+\infty} \|x_k - T(x_k)\|_G^2 = 0$. Property (ii) follows from the observation that if the sequence $\{x_k\}$ has a limit point, then asymptotic regularity implies that it is also a fixed point solution. □

Lemma 1. *If the sequence $\{x_k\}$ has a limit point, all of its limit points are fixed points, and the sequence $\{\|x_k - x^*\|_G\}$ is convergent for every fixed point solution x^*, then the sequence $\{x_k\}$ converges to a fixed point solution.*

Proof. Let $\bar{x}$ be any limit point of the subsequence of x_k (and so a fixed point of the map T). By hypothesis, the sequence $\|x_k - x^*\|_G$ converges to zero, and so the entire sequence $\{x_k\}$ converges to the fixed point solution $\bar{x}$. □

We will use this simple result, together with Property (ii) of Proposition 2, to establish the main convergence result of Section 2.

2. Compatible maps and step sizes: A general case

An adaptive line search framework

One of the main goals of this paper is to determine "good" step sizes a_k, possibly negative, for the general iterative scheme

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

We will achieve this objective by choosing step sizes that are "compatible" with the underlying map T. That is, we will achieve convergence by imposing conditions jointly on the underlying map T and the step sizes a_k.

Motivation:

We will draw our motivation from two results.

1. When the fixed point map T is contractive and the step size $a_k = 1$ at each iteration k, the line search framework coincides with the classical function iteration $x_{k+1} = T(x_k)$ and, as a result, the sequence induced converges to the fixed point solution at rate A. This case imposes a strong condition on the underlying map T.
2. When the fixed point map T is nonexpansive and the step sizes satisfy the conditions of Dunn's theorem, namely, $0 < a_k < 1$ and $\sum_{k=1}^{\infty} a_k(1 - a_k) = +\infty$, then the general line search framework converges (see [12]). This case imposes a strong condition on the step sizes (which slows convergence).

Our analysis will strive to generalize these results by considering the following quantity that simultaneously considers the nature of the underlying fixed point map and the step sizes.

$$A_k(x^*) \equiv \frac{\|x_k - x^*\|_{\bar{G}}^2 - \|T(x_k) - T(x^*)\|_{\bar{G}}^2}{\|x_k - T(x_k)\|_{\bar{G}}^2} + (1 - a_k). \quad (5)$$

In the expression $\bar{G}$ is a positive definite matrix. Notice that when the map T is nonexpansive with Lipschitz constant $0 < A \leq 1$, the numerator of the first term in the definition of A_k^* is at least $(1 - A)\|x_k - x^*\|_{\bar{G}}^2$ and so the first term is nonnegative and therefore for Examples 1 and 2, the quantity $A_k(x^*) > 0$. To permit tradeoffs between the nature of the underlying map and the step sizes, we will impose two assumptions:

Definition 3. *A map T and step sizes a_k are A1-A2 compatible if they satisfy the following two conditions.*

A1. *For any fixed point x^* of the map T, at each iteration k, $a_k A_k(x^*)$ 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.*

We wish to show that if the map T and the step sizes a_k satisfy Assumptions A1 and A2, then the line search scheme $x_{k+1} = x_k + a_k(T(x_k) - x_k)$ converges to a fixed point solution.

Let us first restate the definition of $A_k(x^*)$ in a different, but equivalent, form. Note that since $x^* = T(x^*)$, by adding and subtracting x_k from the first term and expanding, we obtain

$$\|T(x_k) - T(x^*)\|_{\bar{G}}^2 = \|x_k - x^*\|_{\bar{G}}^2 + \|x_k - T(x_k)\|_{\bar{G}}^2 - 2(x_k - T(x_k))^t \bar{G}(x_k - x^*) \quad (6)$$

which when substituted into the definition of $A_k(x^*)$, gives

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

Theorem 1. *Assume that the map T has a fixed point solution x^*. If the map T and step sizes a_k are A1-A2 compatible, then the sequence of iterates $\{x_k\}$ converges to a fixed point solution.*

Proof. Consider any fixed point x^* of the map T. Since $x_{k+1} = x_k + a_k(T(x_k) - x_k)$,

$$\|x_{k+1} - x^*\|_{\bar{G}}^2 = \|x_k - x^*\|_{\bar{G}}^2 - 2a(x_k - T(x_k))^t \bar{G}(x_k - x^*) + a_k^2\|x_k - T(x_k)\|_{\bar{G}}^2. \tag{8}$$

Expression (8) implies that

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

Therefore,

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

Equation (7) implies that

$$\|x_{k+1} - x^*\|_{\bar{G}}^2 = \|x_k - x^*\|_{\bar{G}}^2 - a_k A_k(x^*)\|x_k - T(x_k)\|_{\bar{G}}^2. \tag{9}$$

Relation (9) and assumption A1 imply that the sequence $\|x_k - x^*\|_{\bar{G}}^2$ is nonincreasing and, therefore, is convergent. This result implies that either (a) $\|x_k - T(x_k)\|_{\bar{G}}^2$ converges to zero, or (b) $a_k A_k(x^*)$ converges to zero. Therefore either Proposition 2 (for case (a)) or assumption A2 (for case (b)) implies that every limit point of the sequence of iterates $\{x_k\}$ is also a fixed point solution. Relation (9) implies that $\|x_k - x^*\|_{\bar{G}}^2$ is convergent for every fixed point x^*, and Lemma 1 then implies that the entire sequence of iterates $\{x_k\}$ converges to a fixed point solution. □

This theorem 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 (since from Proposition 1, $A_k(x^*) \geq (1 - A)/(1 + \sqrt{A})^2 > 0$). As we next show, this theorem also includes as special case Dunn's averaging results (see [12], [35]).

Theorem 2. *Suppose that T is a nonexpansive map, the step sizes a_k constitute a Dunn sequence (that is, $\sum_k a_k(1 - a_k) = +\infty$), and the map T has a fixed point solution x^*. Then the map T and the step sizes a_k 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, assumption A1 applies).

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

$$\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 = -a_k A_k(x^*)\|x_k - T(x_k)\|^2, \tag{10}$$

which implies (since $a_k A_k(x^*) \geq 0$) that the sequence $\{\|x_k - x^*\|^2\}$ converges to some constant q for every fixed point solution x^*. Adding and subtracting $T(x_k)$ implies that

$$\begin{aligned} ||x_{k+1} - T(x_{k+1})|| &= ||x_k + a_k(T(x_k) - x_k) - T(x_{k+1})|| \\ &= ||(1 - a_k)(x_k - T(x_k)) + (T(x_k) - T(x_{k+1}))||. \end{aligned}$$

Applying the triangle inequality together with the fact that T is a nonexpansive map implies (since $x_{k+1} - x_k = a(T_k(x_k) - x_k)$) that

$$\|x_{k+1} - T(x_{k+1})\| \leq (1 - a_k)\|x_k - T(x_k)\| + \|x_{k+1} - x_k\| = \|x_k - T(x_k)\|. \tag{11}$$

Therefore, 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, Proposition 2 implies that $\|x_k - T(x_k)\| \geq B > 0$ for all $k \geq k_0$, for a sufficiently large constant k_0. Then by adding the telescoping equations (10) for all $k \geq k_0$ implies that

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

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

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

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 assumptions A1 and A2 and, therefore, Dunn's theorem is a special case of Theorem 1. □

The next example shows that Theorem 1 is a more general result.

Example 1. The map $T(x) = x\sqrt{1 - \|x\|}$ has a unique fixed point $x^* = 0$ over the set $\{x : \|x\| \leq 1\}$ and is nonexpansive around solution x^* (see Definition 1) 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, although nonexpansive, is not a contractive map. Nevertheless, Theorem 1 ensures convergence since $a_k = 1$ for all k is bounded away from zero and $A_k(x^*) = \frac{\|x_k\|}{\|1-\sqrt{1-\|x_k\|}\|^2}$ and so, assumptions A1 and A2 apply. Therefore, Theorem 1 applies for this choice of the map T and step sizes.

Remark. Observe that the steps sizes for the iterates satisfying assumptions A1 and A2 of Theorem 1 need not necessarily lie between zero and one and that the map T need not be nonexpansive. To achieve the joint conditions of assumptions A1 and A2 imposed upon the maps T and step sizes, we might (i) restrict the line searches to a subset of $[0, 1)$ or $(-1, 0]$ permitting maps that satisfy a condition weaker than nonexpansiveness, or (ii) extend the line searches, but as a result, impose stronger assumptions on the map T.

3. Potential optimizing methods

Theorem 1 shows that if we impose joint conditions on the underlying map T and step sizes a_k, the general line search scheme

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

converges. This result provides little guidance, however, concerning the choice of the step sizes. In this section, to develop more operational versions of the general framework, we will choose the step sizes by solving, at each iteration, from a given point x_k of the previous iteration, a one dimensional optimization problem:

$$a_k = \arg\min_{a \in S} g^{x_k}(a), \tag{12}$$

for some step size choice set S and potential function $g^{x_k} : R \to R$.

We consider several choices for the potential function and examine the convergence behavior for the sequence each generates. We also consider specializations when the map T is affine, allowing us to find closed form solutions for the step sizes a_k.

Instead of solving a sequence of optimization problems, we might try to solve a single problem since for any positive definite and symmetric matrix G, the fixed point problem $FP(T, K)$ is equivalent to the minimization problem

$$\min_{x \in K} \|x - T(x)\|_G^2. \tag{13}$$

As is well known (see for example [26]), the difficulty in using this equivalent optimization formulation is that even when T is a contractive map relative to the $\|.\|_G$ norm, the potential function $\|x - T(x)\|_G^2$ need not be convex. The following example illustrates this behavior.

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

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

$x^* = 1$. 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 $[1/2, 9/16)$.

The one-dimensional version of problem (13) (that is, along a line segment) will, in general also be nonconvex, though we can solve it approximately via any one-dimensional search technique.

To implement the framework of Section 2, we will, in general, restrict the choice of potentials.

3.1. Convergence results

Definition 4. *A potential g is* **A1-A2 compatible** *(with respect to the step size search sets) if the step sizes it generates, together with the map T, satisfies conditions A1 and A2.*

From Theorem 1, we immediately have the following result.

Theorem 3. *Assume that the fixed point map T has a fixed point solution x^*. If the potential g is A1-A2 compatible, then the sequence of iterates $\{x_k\}$ generated by the one-dimensional optimization problem (12) converges to a fixed point solution.*

Specializations:

We next consider several specific potential functions. They all are variations of the following two potential functions (with $x(a) = x + a(T(x) - x)$ for a given point x).

1. $G^x(a) = \|x(a) - T(x(a))\|_G^2$.
2. $H^x(a) = [(x(a) - T(x(a)))^t G(x - T(x))]^2$.

The second potential is motivated by the steepest descent method (see Hammond and Magnanti [23]) for solving the unconstrained variational inequality problem,

$$\text{find } x \in R^n \text{ satisfying } f(x) = 0, \qquad VI(f, R^n).$$

Given a point x_k, the steepest descent method sets $x_{k+1} = x_k - a_k f(x_k)$, choosing the step size a_k by solving the following one-dimensional optimization problem

$$\min_a (f(x_k(a))^t G f(x_k))^2,$$

with $x_k(a) = x_k + a(-f(x_k))$. This problem computes the next iterate along the half line originated at iterate x_k in the direction of $-f(x_k)$ so that the vector $f(x_k(a))$ is perpendicular to vector $f(x_k)$ with respect to the matrix G. In the fixed point formulation, we set $f(x) = x - T(x)$.

We consider the following variations of the potentials G^x and H^x:

Scheme	Potential
I	$g_1^x(a) = G^x(a) + \beta\|x(a) - x(0)\|_G^2 = G^x(a) + a^2\beta\|x - T(x)\|_G^2$
II	$g_2^x(a) = G^x(a)$
III	$g_3^x(a) = G^x(a) - \beta\|x(a) - x(0)\|_G^2$
IV	$g_4^x(a) = G^x(a) - \beta\|x(a) - x(0)\|_G\|x(a) - x(1)\|_G$
V	$g_5^x(a) = H^x(a)$
VI	$g_6^x(a) = H^x(a) - \beta[(x(a) - x(0))^t G(x - T(x))][(x(a) - x(1))^t G(x - T(x))]$

In all these schemes, β is a positive constant. We choose the step size a within a search set $S \subseteq R^+$ that is either the entire half line R^+ or is a subset of [0, 1]. For ease of notation, in the follows discussion we will let $G^{x_k}(a) = G^k(a)$, $H^{x_k}(a) = H^k(a)$ and $g_i^{x_k}(a) = g_i^k(a)$, $i = 1, ..., 6$.

By introducing terms of the form $-\beta||x(a) - x(0)||_G^2$, several of these schemes (III, IV and VI) force the iterates $x_{k+1} = x_k(a_k)$ away from the starting point $x_k = x_k(0)$, thus assuring sufficient movement from this point. Two of the schemes (IV and VI) repel the iterate x_{k+1} away from the image $T(x_k) = x_k(1)$ of the starting point under map T. Doing so can ensure convergence for situations when function iteration might cycle.

An additional motivation for introducing Scheme VI comes from trying to remedy a cycling behavior that can occur in Scheme V. The next example illustrates this behavior.

Example 3. When $K = R^n$ and $T(x) = [x_2, -x_1]$, then $x^* = (0, 0)$ is the solution of the fixed point problem $FP(T, R^n)$. Starting from the point $x^0 = (1, 1)$, Scheme V (which coincides with the steepest descent algorithm for solving systems of equations) generates the iterates $x^1 = (1, -1)$, $x^2 = (-1, -1)$, $x^3 = (-1, 1)$, $x^4 = x^0 = (1, 1)$ and, therefore, cycles.

In this example, matrix $M = I - T = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ is positive definite, but $M^2 = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$ is positive semidefinite. This example violates known convergence conditions (see Hammond and Magnanti [23]), namely that the matrices M and M^2 are positive definite. Scheme VI remedies this cycling without requiring that M^2 be positive definite.

The following result summarizes convergence properties for these six schemes.

Theorem 4. *The iterates generated by the special cases of the general averaging framework (Schemes I-VI) converge to a fixed point solution under the following conditions:*

Scheme	Convergence conditions	Step size set S
I	*T contractive, A: $0 \leq \beta < 1 - A$*	$S = R^+$
II	*T contractive (or affine, nonexpansive)*	$S = R^+$ (*or* $S \subseteq (0, 1)$)
III	*T nonexpansive*	$S = [0, c_1] \subseteq [0, 1)$, $0 < c_1 < 1$
IV	*T nonexpansive*	$S = R^+$
V	*T affine, M, M^2 positive definite matrices*	$S = R^+$
VI	*T nonexpansive, $\beta > 4$ (or affine, $\beta > 1$)*	$S = R^+$

To prove this theorem, we show that for each scheme, the corresponding potential is *A1-A2 compatible* so that Theorem 1 applies. We establish these results in Appendix A.

3.2. Convergence rates

Theorem 4 shows that the six special cases of the general averaging framework satisfy assumptions *A1-A2*. The next theorem establishes rates of convergence for five of these schemes.

Theorem 5. *Schemes I-V have the following rates of convergence:*

Scheme	**Convergence rate**
I	$\|x_{k+1} - T(x_{k+1})\|_G^2 \le [A + \beta(1 - a_k^2)]\|x_k - T(x_k)\|_G^2.$
II	$\|x_{k+1} - T(x_{k+1})\|_G^2 \le A\|x_k - T(x_k)\|_G^2.$
III	$\|x_{k+1} - T(x_{k+1})\|_G^2 \le (A - \beta(c_1^2 - a_k^2))\|x_k - T(x_k)\|_G^2,\ 0 < a_k \le c_1 < 1.$
IV	$\|x_{k+1} - T(x_{k+1})\|_G^2 \le [1 - \beta(a_k - \frac{1}{2})^2]\|x_k - T(x_k)\|_G^2.$
V	$\|x_{k+1} - x^*\|_{\bar{G}}^2 \le [1 - \frac{\lambda_{min}((\bar{G})^{-1}\frac{M^2+(M^t)^t}{2})}{\lambda_{max}(\bar{G})}]\|x_k - x^*\|_{\bar{G}}^2.$

Proof.
Scheme I:
First, as we illustrated in Theorem 4, the sequence of iterates $\{x_k\}$ that Scheme I induces converges to the fixed point solution. The iteration of Scheme I implies the inequality

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

Since the map T is a contraction,

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

implying the result. with contractive constant $0 < A + \beta < 1$.

Scheme II:
For nonlinear, contractive mappings T,

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

Scheme III:
The iteration of Scheme III imply that

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

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

Scheme IV:
The iteration of Scheme IV implies that

$$\begin{aligned}&\|x_{k+1} - T(x_{k+1})\|_G^2 - \beta a_k(1 - a_k)\|x_k - T(x_k)\|_G^2 \\ &\quad \leq \|x_k(a^*) - T(x_k(a^*))\|_G^2 - \beta a^*(1 - a^*)\|x_k - T(x_k)\|_G^2,\end{aligned}$$

with $a^* = 1/2$. The nonexpansiveness of map T implies that $\|x_k(a^*) - T(x_k(a^*))\|_G^2 \leq \|x_k - T(x_k)\|_G^2$ and, therefore, setting $h(a) = a(1 - a)$,

$$\begin{aligned}&\|x_{k+1} - T(x_{k+1})\|_G^2 - \beta a_k(1 - a_k)\|x_k - T(x_k)\|_G^2 \\ &\quad \leq \|x_k - T(x_k)\|_G^2 - \beta h(a^*)\|x_k - T(x_k)\|_G^2.\end{aligned}$$

After reordering the terms, we obtain the desired result.
Scheme V:
The proof of this result can be found in [23]. □

3.3. Relationship to variational inequalities

We have considered several special cases of the general line search (averaging) framework. The general framework also includes several other well-known schemes in the literature. As noted in the introduction, in several instances, a variational inequality problem $VI(f, K)$ is equivalent to a fixed point problem $FP(T, K)$ (see for example [11], [20], [42], [53]) and so we can apply our prior results.

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

$$g^x(a) = -f(x(a))^t(T(x(a)) - x(a)) - \frac{1}{2}\|T(x(a)) - x(a)\|_G^2, \tag{14}$$

   Fukushima showed that when f is strongly monotone, the scheme $x_{k+1} = x_k(a_k) = x_k + a_k(T(x_k) - x_k)$, with $a_k \arg\min_{a \in [0,1]} g^k(a)$ computes a variational inequality solution. This scheme and the accompanying convergence result is a special case of our general framework for the choice of potential g given by expression (14). Observe that when $f(x) = G(x - T(x))$, the potential the Fukushima potential becomes $g^x(a) = \|T(x(a)) - x(a)\|_G^2 - \frac{1}{2}\|T(x(a)) - x(a)\|_G^2$ and so is equivalent to optimizing the potential we considered in Scheme II.
2. Wu, Florian and Marcotte [53] have generalized Fukushima's scheme. They considered an operator T that maps a point x onto the unique solution of the problem $\min_{y \in K}\left[f(x)^t(y - x) + \frac{1}{\rho}\phi(x, y)\right]$, for a function $\phi : K \times K \to R$ satisfying 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 solutions of the variational inequality problem $VI(f, K)$ (see [53] for more details). Wu, Florian and Marcotte [53] considered the potential

$$g^x(a) = -f(x(a))^t(T(x(a)) - x(a)) - \frac{1}{\rho}\phi(T(x(a)), x(a)). \tag{15}$$

Notice as in the previous case, this scheme becomes a special case of our general framework for the choice of potential g^x given in (15).

The convergence of these schemes follows from Theorem 1. Observe that when f is strongly monotone (which the developers of these schemes impose), assumptions A1 and A2 hold. These assumptions are valid because (i) the quantity $A_k(x^*)$ is bounded away from zero, and (ii) when $f(x) = x - T(x)$ and function f is strongly monotone then, for the function ϕ as defined above (and for the particular Fukushima case, $\phi(x, y) = \|x - y\|_G^2$), it is possible to show that the potentials (14) and (15) "repel" step sizes a_k from zero unless the current iterate is a solution.

3.4. Lessons

The analysis of these schemes leads to the following conclusions:

1. When the fixed point map T is affine, except for Scheme V, the schemes converge for nonexpansive maps.
2. When the fixed point map T is affine, we need not impose any restrictions on the step sizes for Schemes I, II and V (provided that map T satisfies some additional properties).
3. If the scheme repels the iterates from both boundaries x_k and $T(x_k)$ (as in Schemes IV and VI), then we need not impose restrictions on the step size search set S, but rather we can allow it to be the entire half line R^+.
4. If the scheme does not repel the iterates from both boundaries, then we either need to impose stronger conditions on the fixed point map T (such as contractiveness) or restrict the step size search set S within a subset of (0, 1).

4. Optimal rates of convergence

In this section, we examine the following question: *when is the rate of convergence for the general adaptive line search framework and its special cases better than that of function iteration*?

For contractive maps, the function iteration estimate $\|x_{k+1} - x^*\|_G^2 = \|T(x_k) - T(x^*)\|_G^2 \leq A\|x_k - x^*\|_G^2$, provides a guaranteed convergence rate of A. We will show that if the step sizes generated by the line search procedure in the general adaptive framework lie within a certain range, then the adaptive line search procedure yields a better guaranteed rate of convergence. We also show that the six schemes we have introduced give rise to such step sizes.

As an example, from the starting point of $x_k = (0, 1)$ for the map $T(y, z) = b(z, -y)$, then $x_{k+1} = x_k + a_k(T(x_k) - x_k) = (a_k b, 1 - a_k)$ and

$$\|x_{k+1} - x^*\|^2 = \|x_{k+1} - 0\|^2 = \|(a_k b, 1 - a_k)\|^2 = a_k^2 b^2 + a_k^2 + 1 - 2a_k.$$

Therefore, the step size $a_k = \frac{1}{1+b^2}$ minimizes $\|x_{k+1} - 0\|^2$ and the norm minimizing step size ranges from $a_k = \frac{1}{2}$ when $b = 1$ to $a_k = 1$ as b approaches 0. The analysis to follow will generalize this observation.

Theorem 6. *For contractive maps with constant A, if the quantities $(a_k - 1)(A_k(x^*) - 1)$, for all k, are nonnegative then the adaptive line search framework has at least as good a rate of convergence as function iteration. If the quantities $(a_k - 1)(A_k(x^*) - 1)$ are bounded away from zero (that is, for all k, $(a_k - 1)(A_k(x^*) - 1) \geq q^2$, for some constant $q > 0$), then the adaptive line search framework has a better guaranteed rate of convergence than function iteration.*

Proof. The definition $x_{k+1} = x_k + a_k(T(x_k) - x_k)$, the expansion

$$\|x_{k+1} - x^*\|_{\bar{G}}^2 = \|x_k - x^*\|_{\bar{G}}^2 + a_k^2 \|x_k - T(x_k)\|_{\bar{G}}^2 - 2a_k(x_k - T(x_k))^t \bar{G}(x_k - x^*), \quad (16)$$

and a substitution for $\|x_{k+1} - x^*\|_{\bar{G}}^2$ from (6) implies that

$$\begin{aligned}\|x_{k+1} - x^*\|_{\bar{G}}^2 = \|T(x_k) - T(x^*)\|_{\bar{G}}^2 + (a_k^2 - 1)\|x_k - T(x_k)\|_{\bar{G}}^2 \\ -2(a_k - 1)(x_k - T(x_k))^t \bar{G}(x_k - x^*).\end{aligned} \quad (17)$$

Substituting for $A_k(x^*)$ from expression (7) gives

$$\begin{aligned}\|x_{k+1} - x^*\|_{\bar{G}}^2 &= \|T(x_k) - T(x^*)\|_{\bar{G}}^2 + [(a_k^2 - 1) - (a_k - 1)(A_k(x^*) + a_k)] \\ &\quad \times \|x_k - T(x_k)\|_{\bar{G}}^2 \\ &= \|T(x_k) - T(x^*)\|_{\bar{G}}^2 - [(a_k - 1)(A_k(x^*) - 1)] \\ &\quad \times \|x_k - T(x_k)\|_{\bar{G}}^2.\end{aligned} \quad (18)$$

Using the fact that T is contractive, we obtain

$$\|x_{k+1} - x^*\|_{\bar{G}}^2 = A\|x_k - x^*\|_{\bar{G}}^2 - (a_k - 1)(A_k(x^*) - 1)\|x_k - T(x_k)\|_{\bar{G}}^2. \quad (19)$$

Consequently, if $(a_k - 1)(A_k(x^*) - 1) \geq 0$ then Proposition 1 implies that

$$\|x_{k+1} - x^*\|_{\bar{G}}^2 \leq [A - (a_k - 1)(A_k(x^*) - 1)(1 - \sqrt{A})]\|x_k - x^*\|_{\bar{G}}^2. \quad (20)$$

As a result, whenever when $(a_k - 1)(A_k(x^*) - 1) \geq 0$, the guaranteed rate of convergence for adaptive line search is at least as good a that for function iteration, and when $(a_k - 1)(A_k(x^*) - 1) \geq q^2$ for some constant $q > 0$, then adaptive line search has a better rate of convergence than function iteration. □

The similarity between the term $(a_k - 1)(A_k(x^*) - 1)$ that arises in this result and the term $a_k A_k(x^*)$ that plays such a central role in analyzing the general adaptive line search framework is striking.

We obtain the best possible improvement with the largest possible value of the quantity $(a_k - 1)(A_k(x^*) - 1)$. We will refer to any step size giving this maximum value as an "optimal" step size.

The previous theorem shows that for certain types of maps and for a certain range of step sizes, the adaptive line search framework provides a better rate of convergence than the classical function iteration $x_{k+1} = T(x_k)$. The next result provides several examples.

Proposition 3. *For suitable choice of step sizes a_k, the condition $(a_k - 1)(A_k(x^*) - 1) \geq q^2 > 0$ is valid in any of the following circumstances.*

1. *The map T contractive, but with a contraction constant close to 1, that is, for some constants $A \in (0, 1)]$ and $\bar{A} \in (0, 1)$,*
$$A\|x - x^*\|_{\bar{G}}^2 \geq \|T(x) - T(x^*)\|_{\bar{G}}^2 \geq \|x - x^*\|_{\bar{G}}^2 - \bar{A}\|x - T(x)\|_{\bar{G}}^2.$$
2. *The map T is a tightly nonexpansive map, that is, $\|T(x) - T(x^*)\|_{\bar{G}}^2 \approx \|x - x^*\|_{\bar{G}}^2$.*
3. *The map T is firmly contractive, that is, $\|T(x) - T(x^*)\|_{\bar{G}}^2 \leq A\|x - x^*\|_{\bar{G}}^2 - \|x - T(x) - x^* + T(x^*)\|_{\bar{G}}^2$, for some constant $A \in (0, 1)$.*

Proof. 1. Using expression (6) to eliminate the first two terms from the inequality
$$\|T(x_k) - T(x^*)\|_{\bar{G}}^2 \geq \|x_k - x^*\|_{\bar{G}}^2 - \bar{A}\|x_k - T(x_k)\|_{\bar{G}}^2$$
gives
$$\|x_k - T(x_k)\|_{\bar{G}}^2 - 2(x_k - T(x_k))^t \bar{G}(x_k - T(x_k)) \geq -\bar{A}\|x - T(x)\|_{\bar{G}}^2.$$
Consequently, expression (7) implies that $A_k(x^*) + a_k \leq 1 - \bar{A}$, from which we conclude that $(a_k - 1)(A_k(x^*) - 1) \geq (a_k - 1)(-a_k - \bar{A})$, whenever $a_k < 1$. The choice $a_k = \frac{1-\bar{A}}{2}$ maximizes the righthand side of the last inequality and so is an optimal step size. For this choice of step size, $(a_k - 1)(A_k(x^*) - 1) \geq \frac{1}{4}(\bar{A} + 1)(\bar{A} + 2)$.

2. If the map T is tightly nonexpansive, then $\bar{A}$ in case 1. is approximately 0. Similarly to case 1., the optimal step size is approximately $\frac{1}{2}$, and so $(a_k - 1)(A_k(x^*) - 1) \geq \frac{1}{4}$.

3. For the firmly contractive map T, expression (6) implies that
$$(x - T(x) - x^* + T(x^*))^t \bar{G}(x - x^*) \geq \frac{1 - A}{2}\|x - x^*\|_{\bar{G}}^2 + \|x - T(x)\|_{\bar{G}}^2,$$
and Proposition 1 implies that
$$(x - T(x) - x^* + T(x^*))^t \bar{G}(x - x^*) \geq \frac{1 - A}{2(1 + \sqrt{A})^2}\|x - T(x)\|_{\bar{G}}^2 + \|x - T(x)\|_{\bar{G}}^2.$$
But then from expression (7), $A_k(x^*) + a_k \geq \frac{1-A}{(1+\sqrt{A})^2} + 2$ and so, if $a_k > 1$, then $(a_k - 1)(A_k(x^*) - 1) \geq (a_k - 1)\left(\frac{1-A}{(1+\sqrt{A})^2} + 1 - a_k\right)$. The step size $a_k = 1 + \left(\frac{1-A}{2(1+\sqrt{A})^2}\right)$ maximizes the righthand side of the previous inequality, specifying an optimal step size greater than 1.

□

Suppose the underlying fixed point map is affine, that is, $x - T(x) = Mx$ for some matrix M. Then the step sizes in the six schemes we introduced in the previous section always satisfy the condition of Theorem 6. As a result, these schemes (I–V) have better rates of convergence than the classical function iteration. In particular, this condition holds in the following situations.

- The map T is firmly contractive and $\beta < (1 - A)/2$ for Scheme I.
- The map T is contractive, and the matrix $M^t M - M$ or the matrix $M - M^t M$ is strongly positive definite (i.e., for some $C > 0$, for all $x \in R^n$, $x^t(M^t M - M)x \geq Cx^t x$ or $x^t(M - M^t M)x \geq Cx^t x$) for Scheme II.
- The map T is contractive for Scheme III.
- The map T is firmly contractive and $\beta < 1 - A$ for Scheme IV.
- The map T is contractive and M is a symmetric matrix, and the matrix $M - I$ or the matrix $I - M$ is strongly positive definite for Scheme V.

We establish these results in Appendix B. We also show that the guaranteed rates of convergence for Schemes II and V are always at least as good as function iteration even without the additional conditions imposed upon the matrix M.

This discussion shows that for certain fixed point maps T, the six schemes exhibit better rates of convergence than the classical function iteration.

5. Inexact line searches

One issue arises naturally when attempting to implement any of the methods examined in Section 3: *suppose we perform line searches inexactly (especially in the general nonlinear case), will the various schemes still converge?*

We can determine step sizes at each step by applying an Armijo-type rule of the following form. Let g_i denotes the potential function we have used previously for scheme i and d_k denote $T(x_k) - x_k$. For positive constants $D > 0$ and $0 < b < 1$, we wish to find the smallest nonnegative integer l_k so that a step length $a_k = b^{l_k}$ in the appropriate line search set S that satisfies the condition

$$g_i^k(a_k) - g_i^k(0) \leq -Db^{l_k} \|d_k\|_G^2. \tag{21}$$

This condition aims to ensure that the potential function decreases by a sufficient amount at each iteration. We next establish convergence of the Armijo-type inexact line searches (21) for the potentials of Schemes II and IV that we considered in Section 3 if we impose appropriate conditions on the fixed point map T. We first consider Scheme II.

Theorem 7. *Suppose that in a fixed point problem $FP(T, K)$, T is a contractive map relative to the $\|.\|_G$ norm with a Lipschitz constant $A \in (0, 1)$. Then with a choice of $D \leq 1 - \sqrt{A}$ in an Armijo-type search (21), the potential minimization scheme with a potential $g_2^k(a) = \|x_k(a) - T(x_k(a))\|_G^2$ and step sizes in the set $S = [0, 1]$ generates a sequence that converges to a fixed point solution.*

Proof. We first show that the Armijo-type line search (21) has a solution. Adding and subtracting $T(x_k)$ implies that

$$\begin{aligned}||x_k(a) - T(x_k(a))||_G &= ||x_k + a(T(x_k) - x_k) - T(x_k(a))||_G \\ &= ||(1 - a)(x_k - T(x_k)) + (T(x_k) - T(x_k(a)))||_G.\end{aligned}$$

Applying the triangle inequality together with the fact that T is a contractive map with constant A, further implies that

$$\begin{aligned}\|x_k(a) - T(x_k(a))\|_G &\leq (1-a)\|x_k - T(x_k)\|_G + \sqrt{A}\|x_k(a) - x_k\|_G \\ &= (1 - a(1-\sqrt{A}))\|x_k - T(x_k)\|_G.\end{aligned}$$

Consequently, for all $0 \leq a \leq 1$,

$$\begin{aligned}\|x_k(a) - T(x_k(a))\|_G^2 &\leq [1 - a(1-\sqrt{A})]^2\|x_k - T(x_k)\|_G^2 \\ &\leq [1 - a(1-\sqrt{A})]\|x_k - T(x_k)\|_G^2.\end{aligned}$$

Therefore, $g_2^k(a) - g_2^k(0) \leq -(1-\sqrt{A})a\|d_k\|_G^2$ and so if we choose $D \leq 1 - \sqrt{A}$ and $a = b^{l_k}$ with $l_k = 1$, then the constants D, and l_k satisfy the Armijo-type rule (21) for any choice of $0 < b < 1$.

Consequently,

$$\|x_k - T(x_k)\|_G^2 - \|x_{k+1} - T(x_{k+1})\|_G^2 \geq b(1-\sqrt{A})\|x_k - T(x_k)\|_G^2. \qquad (22)$$

and so

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

and, as a result, the sequence $\|x_k - T(x_k)\|_G^2$ converges to zero. Proposition 1 implies that the entire sequence $\{x_k\}$ converges to a fixed point solution. □

Theorem 8. *Consider the fixed point problem $FP(T, K)$ with nonexpansive maps T relative to the $\|.\|_G$ norm. Suppose the fixed point problem has a solution. If we apply an Armijo-type (21) line search to the potential function g_4^k in Scheme IV, with step sizes $a \in S = [0, 1]$, then the sequence of iterates x_k that this line search generates converges to a fixed point solution.*

Proof. We will first show that the Armijo-type (21) line search havs a solution and then that when the step sizes satisfy the Armijo inequality, the iterates converge to a fixed point solution.

For Scheme IV, we can rewrite the potential $g_4^k(a)$ as

$$g_4^k(a) = \|x_k(a) - T(x_k(a)\|_G^2 - \beta h_4(a)\|x_k - T(x_k)\|_G^2,$$

with $h_4(a) = a(1-a)$. As shown in expression (11), for nonexpansive maps T,

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

for all $0 \leq a \leq 1$, implying that

$$g_4^k(a) - g_4^k(0) \leq -\beta h_4(a)\|x_k - T(x_k)\|_G^2, \qquad (23)$$

for all $0 \leq a \leq 1$.

Therefore, the step sizes a_k will satisfy the Armijo-type line search condition (21) whenever $\beta h_4(a_k) \geq D a_k$.

For the potential $g_4^k(a)$, the condition $\beta h_4(a_k) \geq D a_k$ becomes $\beta a_k(a_k - 1) \geq D a_k$ and so $a_k \leq 1 - D/\beta$. The step size $a_k = b^{l_k}$ satisfies this condition whenever $l_k \geq$

$\frac{\ln(1-D/\beta)}{\ln b} = q$. Any value of l_k satisfying this inequality satisfies the Armijo step size condition and so the smallest value of l_k satisfying Armijo-type inequality (21) is no more than $\lceil q \rceil$, the integer round up of q.

We have shown that for potential $g_4^k(a)$, the Armijo-type inequality (21) has a solution. To establish convergence, we will use Theorem 1. Since the map T is nonexpansive, $A_k(x^*) \geq (1-a_k) \geq 0$ and so by Theorem 1, $a_k A_k(x^*) \geq a_k(1-a_k)$. For any $b < 1$ and $l_k \leq \lceil q \rceil$, $b^{\lceil q \rceil} \leq a_k = b^{l_k} \leq b$ and so $a_k(1-a_k) \geq b^{\lceil q \rceil}(1-b) > 0$. Therefore, $a_k A_k(x^*)$ is bounded away from zero and so by Theorem 1, the iterates converge to a fixed point solution. □

6. Conclusions-open questions

We have introduced and studied adaptive line search methods for solving fixed point and variational inequality problems. Motivated by a desire to integrate, in a single result, the convergence behavior of classical Banach functional iteration and several averaging methods, we first established a convergence framework that imposes conditions jointly on the underlying fixed point (or variational inequality) map and the iterative step sizes. Adaptive line search methods that satisfy the joint condition are able to compute fixed points when the underlying maps satisfy properties weaker than contractiveness. As a specific instantiation of the general framework, we considered a general scheme for determining step sizes dynamically by optimizing potential functions. We considered several choices of potential functions that optimize, in some sense, the distance between the current iterate and the image of the fixed point mapping of the current iterate. We established convergence and convergence rates for these choices of potential functions. We also showed that under appropriate conditions the general methods we have considered have a better rate of convergence than function iteration. Since the line searches we proposed might be hard to perform exactly, we also considered inexact line searches.

The results in this paper apply to more general potentials of the type

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

In this more general setting, the function $h : R \to R^+$ might be, for example, $h(a) = a^2$ or $h(a) = a(1-a)$ while function $P : R^n \times R^n \to R^+$ generalizes the notion of a norm. Examples of P include $P(x, y) = \|x-y\|_G^2$ and $P(x, y) = (f(x) - f(y))^t(x-y)$.

Several open questions arise naturally from our analysis:

- *How does the behavior of the schemes we introduced compare for various choices of potentials in practice?*
- *It might be preferable to consider averages of the type* $x_k = \frac{a^1 x_1 + a^2 T(x_1) + \cdots + a^k T(x_{k-1})}{a^1 + \cdots + a^k}$, *and optimize potentials involving all* a^i*'s, instead of using* $x_k(a) = x_k + a_k(T(x_k) - x_k)$ *and optimize potentials involving only* $a_k = \frac{a^k}{a^1 + \cdots + a^k}$.
- *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?*

In closing, we note that some computational experience reported by Bottom [3] and Bottom et al. [4], though hardly conclusive, provides limited evidence concerning the potential computational benefits of the schemes we have introduced. These researchers have used the adaptive line search framework introduced in this paper to solve an anticipatory route guidance problem in the area of Intelligent Transportation Systems, using the Boston Central Artery network during the morning traffic period as a test example. Their empirical results showed that the adaptive averaging framework outperforms averaging with a step size $a_k = 1/k$ at each iteration. Scheme IV proved to be slightly more effective than using line searches with constant step sizes. Functional iteration did not converge for these test problems. More extensive computational testing would be required to properly access the efficiency of the methods we have examined and might even suggest useful enhancements of these methods.

Acknowledgements. We are grateful to the associate editor and the referees for their insightful comments and suggestions that have helped us improve both the exposition and the content of this paper.

References

1. Baillon, J.B.: 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 (1975)
2. Bertsekas, D.P.: Nonlinear programming. Athena Scientific, Belmont, MA (1996)
3. Bottom, J.(1998): Generation of Anticipatory Route Guidance, presentation in TRISTAN III, Triennial Symposium on Transportation Analysis, San Juan, Puerto Rico, also in Optimization Days, Montreal, May 1999
4. Bottom, J., Ben-Akiva, M., Bierlaire, M., Chabini, I., Koutsopoulos, H., Yang, Q.: Investigation of route guidance generation issues by simulation with DynaMIT", to appear in Proceedings of 14th International Symposium on Transportation and Traffic Theory, Jerusalem (1999)
5. Browder, F.E.: Convergence of approximants to fixed points of nonexpansive nonlinear mappings in Banach spaces. Arch. Rational Mech. Anal. **24**, 82–90 (1967)
6. Browder, F.E., Petryshn, W.V.: Construction of fixed points of nonlinear mappings in a Hilbert space. Journal of Analysis and Applications **20**, 197–228 (1967)
7. Bruck, Jr. R.E.: 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 (1973)
8. Bruck, Jr. R.E. : 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 (1977)
9. Calamai, P.H., More, J.J.: Projected gradient methods for linearly constrained problems. Mathematical Programming **39**, 93–116 (1987)
10. Christodouleas, J.: Applications of fixed point problems in scheduling. Private communication (1996)
11. Dafermos, S.C.: An iterative scheme for variational inequalities. Mathematical Programming **26**, 40–47 (1983)
12. Dunn, J.C.: On recursive averaging processes and Hilbert space extensions of the contraction mapping principle. Journal of the Franklin Institute **295**, 117–133 (1973)
13. Dunn, J.C.: Convergence rates for conditional gradient sequences generated by implicit step length rules. SIAM Journal on Control and Optimization **18** (5), 473–487 (1980)
14. Dunn, J.C.: Newton's method and the Goldstein step length rue for constrained minimization problems. SIAM Journal on Control and Optimization **18**, 659–674 (1980)
15. Dunn, J.C.: Global and asymptotic convergence rate estimates for a class of projected gradient processes. SIAM Journal on Control and Optimization **19** (3), 368–400 (1981)
16. Dunn, J.C., Harshbarger, S.: Conditional gradient algorithms with open loop step size rules. Journal of Mathematical Analysis and Applications **62**, 432–444 (1978)
17. Eckstein, J., Bertsekas, D. P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical Programming **55**, 293–318 (1992)
18. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Volumes I and II, Springer Verlag, Berlin (2003)

19. Florian, M., Hearn, D.: Network Equilibria. Handbook of Operations Research and Management Science. Volume Network Routing (M. Ball, T. Magnanti, C. Monma and G. Nemhauser eds.), Elsevier (1995)
20. M. Fukushima: Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. Mathematical Programming **53** (1), 99–110 (1992)
21. Gabay, D.: 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 (1983)
22. Halpern, B.: Fixed points of nonexpansive maps. Bulletin of the American Mathematical Society **73**, 957–961 (1967)
23. Hammond, J.H., Magnanti, T.L.: Generalized descent methods for asymmetric systems of equations. Mathematics of Operations Research **12**, 678–699 (1987)
24. Harker, P.T.: 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 (1993)
25. Harker, P.T., Pang, J.S.: Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. Mathematical Programming **48**, 161–220 (1990)
26. Hoang, T.: Computing fixed points by global optimization methods, in Fixed point theory and applications M.A. Thera and J-B Baillon Eds. 231–244 (1991)
27. Ishikawa, S.: Fixed points and iteration of a nonexpansive mapping in a Banach space. Proceedings of the American Mathematical Society **59** (1), 65–71 (1976)
28. Itoh, T., Fukushima, M., Ibaraki, T.: An iterative method for variational inequalities with application to traffic equilibrium problems. Journal of the Operations Research Society of Japan **31** (1), 82–103 (1988)
29. Kaniel, S.: Construction of a fixed point for contractions in banach space. Israel J. Math. **9**, 535–540 (1971)
30. Kinderlehrer, D., Stampacchia, G.: An introduction to variational inequalities and their application. Academic Press, New York (1980)
31. Knopp, K.: Theory and application of infinite series. Hafner Publishing Company N.Y. (1947)
32. Kolmogorov, A.N., Fomin, S.V.: Introduction to real analysis. Dover Publications, New York (1970)
33. Magnanti, T.L., Perakis, G.: A unifying geometric solution framework and complexity analysis for variational inequalities. Mathematical Programming **71**, 327–351 (1995)
34. Magnanti, T.L., Perakis, G.: The orthogonality theorem and the strong-f-monotonicity condition for variational inequality algorithms. SIAM Journal of Optimization **7** (1), 248–273 (1997)
35. Magnanti, T.L., Perakis, G.: Averaging schemes for variational inequalities and systems of equations. Mathematics of Operations Research **22** (3), 568–587 (1997)
36. Mann, W.R.: On recursive averaging processes and Hilbert space extensions of a contraction mapping principle. Proceedings of the American Mathematical Society **4**, 506–510 (1953)
37. Marcotte, P., Dussault, J-P: A note on a globally convergent Newton method for solving monotone variational inequalities. Operations Research Letters **6** (1), 35–42 (1987)
38. Marcotte, P., Wu, J.H.: On the convergence of projection methods: application to the decomposition of affine variational inequalities. JOTA **85** (2), 347–362 (1995)
39. Nagurney, A.: Network Economics. Kluwer Academic Publishers, Boston (1993)
40. Opial, Z.: Weak convergence of the successive approximation for non-expansive mappings in Banach Spaces. Bulletin A.M.S. **73**, 123–135 (1967)
41. Ortega, J.M., Rheinboldt, W.C.: Iterative solution of nonlinear equations in several variables. Academic Press , New York (1970)
42. Pang, J.S., Chan, D.: Iterative methods for variational and complementarity problems. Mathematical Programming **24**, 284–313 (1982)
43. Pang, J.S.: Complementarity problems. Handbook of Global Optimization, eds. R. Horst and P. Pardalos, 271–329, Kluwer Academic Publishers. (1994)
44. Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. Journal of Mathematical Analysis and Applications **72**, 383–390 (1979)
45. Rockafellar, R.T.: Monotone operators and the proximal algorithm. SIAM Journal on Control and Optimization **14**, 877–898 (1976)
46. Sibony, M.: Methodes iteratives pour les equations et inequations aux derivees partielles nonlinearies de type monotone. Calcolo **7**, 65–183 (1970)
47. Sine, R.C.: Fixed Points and Nonexpansive Mappings. Contemporary Mathematics, **18**, American Mathematical Society (1983)
48. Sun, J.: An affine scaling method for linearly constrained convex programs. Preprint, Department of Industrial Engineering and Management Sciences, Northwestern University (1990)
49. Taji, K., Fukushima, M., Ibaraki, T.: A globally convergent Newton method for solving strongly monotone variational inequalities. Mathematical Programming **58** (3), 369–383 (1993)

50. Todd, M.J.: The computation of fixed points and applications. Lecture Notes in Economics and Mathematical Systems **124**, Springer-Verlag (1976)
51. Tseng, P.: Further applications of a matrix splitting algorithm to decomposition in variational inequalities and convex programming. Mathematical Programming **48**, 249–264 (1990)
52. Tseng, P., Bertsekas, D.P., Tsitsiklis, J.N.: Partially asynchronous, parallel algorithms for network flow problems and other problems. SIAM Journal on Control and Optimization **28** (3), 678–710 (1990)
53. Wu, J.H., Florian, M., Marcotte, P.: A general descent framework for the monotone variational inequality problem. Mathematical Programming **61** (3), 281–300 (1993)
54. Zhu, D.L., Marcotte, P.: Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. SIAM Journal on Optimization **3** (6), 714–726(1996)
55. Zhu, D.L., Marcotte, P.: Modified descent methods for solving the monotone variational inequality problem. Operations Research Letters **14** (2), 111–120 (1993)

Appendix A - Convergence of the six schemes

To establish the convergence of the averaging framework for each of the Schemes I–VI, we will show that the potential function for each case is A1-A2 compatible and, therefore, Theorem 3 applies.

Scheme I:

We first establish the following lemma.

Lemma 1. *Suppose T is a contractive map relative to the $\|.\|_G$ norm and $\beta < 1 - A$. Then if $\lim_{k\to+\infty} a_k = 0$, the sequence $\{x_k\}$ converges to the unique solution (that is, assumption A2 is valid).*

Proof. Suppose $\|x_k - T(x_k)\|_G^2 > 0$. 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 $\beta < 1 - A$)

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

This contradiction implies that $\lim_{k\to+\infty} \|x_k - T(x_k)\|_G^2 = 0$, and then Proposition 2 implies that the sequence $\{x_k\}$ converges to the unique fixed point solution. □

If we restrict the line searches so that $a_k \in [0, 1]$, assumption A1 follows from the contractive property of the map T so Lemma 1 implies that for any choice $\beta < 1 - A$, the step sizes a_k satisfy assumptions A1 and A2.

To develop an alternative proof of convergence for Scheme I, we could perform unrestricted line searches, choosing step sizes $a_k \in R^+$. Then the general iteration of Scheme I implies that

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

and so

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

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

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 1 implies that for any choice $\beta < 1 - A$, the sequence $\{x_k\}$ converges to a solution. In case (ii), Proposition 2 and the fact that map T is contractive imply that the sequence $\{x_k\}$ converges to the solution.

Scheme II:

To establish assumptions *A1-A2*, we establish the following lemma, demonstrating the validity of assumption A2.

Lemma 2. *If T is a contraction with contraction constant A, then $\lim_{k\to\infty} a_k = 0$ implies that the sequence of iterates $\{x_k\}$ converges to the unique fixed point solution.*

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, $\|T(x_k) - T(T(x_k))\|_G^2 \leq A\|x_k - T(x_k)\|_G^2$ 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 2 together with the contractiveness of map T (implying uniqueness of solution for the fixed point problem) implies the conclusion. □

Lemma 2 showed that the step sizes a_k satisfy assumption A2. Moreover,

1. If T is a *contractive* map relative to the $\|.\|_G$ norm and we restrict the line search to $[0, 1]$, then assumptions A1 and A2 follow from Theorem (3).
2. In the *affine, nonexpansive* case when $I - T = M$ for some matrix M, we do not need to restrict the line search. In fact, since $\|T(d)\|_G^2 = \|(I - M)d\|_G^2 \leq \|d\|_G^2$ is equivalent to $d^t Md \geq \frac{1}{2}\|Md\|_G^2$, $a_k = \frac{d_k^t GM(d_k)}{\|Md_k\|_G^2} \geq \frac{1}{2}$, 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 and A2 are valid.

Alternatively, for nonlinear, contractive mappings, the convergence of the iterates of Scheme II 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 2 and the contractiveness of map T then imply convergence to a fixed point solution. □

Scheme III:

Lemma 3. *If T is a nonexpansive map relative to the $\|.\|_G$ norm, then assumptions A1-A2 are valid.*

Proof. Observe that in Scheme III, $g_3^k(a_k) = G^k(a_k) - \beta\|x_k(a_k) - x_k(0)\|_G^2, a_k \in [0, c_1]$. In addition, if $c_1 < 1$, then 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, assumptions A1 and A2 are valid.

On the other hand, if $c_1 = 1$, then we need to assume that T is a contractive map, that is, $0 < A < 1$. Assumptions A1 and A2 are valid because $A_k(x^*) \geq (1 - A)\frac{\|x_k - x^*\|_G^2}{\|x_k - T(x_k)\|_G^2} > 0$. □

Scheme IV:
The next lemma shows that assumptions *A1-A2* are valid.

Lemma 4. *Consider a nonexpansive map T relative to the $\|.\|_G$ norm. If $\lim_{k\to+\infty} a_k = 1$, then every limit point of the sequence of iterates $\{x_k\}$ is a fixed point solution. Then assumptions A1 and A2 are valid.*

Proof. If we choose step sizes $a_k \in [0, 1]$, then assumption A1 is valid. Moreover, since

$$\begin{aligned}\|x_{k+1} - T(x_{k+1})\|_G^2 &\le \|x_k(\frac{1}{2}) - T(x_k(\frac{1}{2}))\|_G^2 - \beta(a_k - \frac{1}{2})^2\|x_k - T(x_k)\|_G^2 \\ &\le [1 - \beta(a_k - \frac{1}{2})^2]\|x_k - T(x_k)\|_G^2,\end{aligned}$$

(we establish this result as part of Theorem 4), 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 $k \ge k_0$, $A_k(x^*) \ge 1 - a_k \ge 0$, that is, assumption A2 applies. □

Scheme V:
To prove assumptions *A1-A2*, we first establish the following lemma.

Lemma 5. *If T is a contraction map, then $\lim_{k\to+\infty} a_k = 0$ implies that the sequence of iterates $\{x_k\}$ converges to the fixed point solution.*

Proof. Assume T is a contraction with contractive constant $A \in (0, 1)$ and so the fixed point problem has a unique solution. If $\lim_{k\to+\infty} a_k = 0$, then $\lim_{k\to+\infty}[(x_k - T(x_k))^t(x_k - T(x_k))]^2 \le \lim_{k\to+\infty}[(T(x_k) - T^2(x_k))^t(x_k - T(x_k))]^2$. This result implies that $\lim_{k\to+\infty}\|x_k - T(x_k)\|^4 \le \lim_{k\to+\infty} A^2\|x_k - T(x_k)\|^4$, but since $A < 1$, this conclusion is a contradiction unless $\lim_{k\to+\infty}\|x_k - T(x_k)\|_G = 0$. □

Since $a_k = \frac{d_k^t d_k}{d_k^t M d_k}$, assumption A1 is valid when M^2 and M are positive definite matrices. In particular, if 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,

$$A_k(x^*) = \frac{(x_k - x^*)^t M^2(x_k - x^*)}{d_k^t M d_k} \ge \frac{\lambda_{min}(\frac{M^2+(M^2)^t}{2})}{\lambda_{max}(M^t \bar{G} M)} = c > 0,$$

(for $x \ne x^*$). This relation implies assumptions *A1-A2*. □

Scheme VI:
To establish assumptions *A1-A2*, we first establish the following lemma.

Lemma 6. *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

$$\lim_{k\to+\infty} g_6^k(0) = \lim_{k\to+\infty} \|d_k\|_G^4 \leq \lim_{k\to+\infty} g_6(a) \leq [1-\beta a(1-a)] \lim_{k\to+\infty} \|d_k\|_G^4. \quad (24)$$

If $\lim_{k\to+\infty} d_k \neq 0$, then for all $a \in (0,1)$, $\lim_{k\to+\infty} g_6^k(0) > [1-\beta a(1-a)]\|d_k\|_G^4$, 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

(i) If T is an affine map and if we choose $\beta > 1$, then every limit point of the sequence $\{x_k\}$ is a fixed point solution.

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

This result implies that $\|d_k\|_G$ converges to zero. Therefore, Proposition 2 implies that every limit point of the sequence $\{x_k\}$ is a fixed point solution.

(ii) If T is a nonlinear, nonexpansive map, and $\beta > 4$, then every limit point of the sequence $\{x_k\}$ is a fixed point solution.

Suppose that a_k converges to one, $\beta > 4$, and no limit point of the sequence $\{x_k\}$ is a fixed point solution. For ease of notation, let $P^{x_k}(a) = [(x_k(a) - T(x_k(a)))^t G(x_k - T(x_k))]$, with $x_k(a) = x_k + a(T(x_k) - x_k)$. If no limit point of the sequence x_k is a fixed point solution, then $\lim_{k\to+\infty} P^{x_k}(0) = \lim_{k\to+\infty} ||x_k - T(x_k)||_G^2 \neq 0$. Notice that

$$\lim_{k\to+\infty} g_6^k(a_k) = \lim_{k\to+\infty} P^{x_k}(a_k)^2 = \lim_{k\to+\infty} P^{x_k}(0)^2 > 0.$$

For $\beta > 4$, we can choose an $\bar{a}$ so that $\frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{\beta}} < \bar{a} < \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{\beta}}$. which implies that $1 - \beta\bar{a}(1-\bar{a}) < 0$. Furthermore, since T is a nonexpansive map, Cauchy's inequality implies that,

$$\begin{aligned} P^{x_k}(\bar{a})^2 &= [(x_k(\bar{a}) - T(x_k(\bar{a})))^t G(x_k - T(x_k))]^2 \\ &\leq \|x_k(\bar{a}) - T(x_k(\bar{a}))\|_G^2 \|x_k - T(x_k))\|_G^2 \leq \|x_k - T(x_k))\|_G^4 = P^{x_k}(0)^2. \end{aligned}$$

Notice that

$$\begin{aligned} \lim_{k\to+\infty} g_6^k(\bar{a}) &= \lim_{k\to+\infty} P^{x_k}(\bar{a})^2 - \beta\bar{a}(1-\bar{a}) \lim_{k\to+\infty} P^{x_k}(0)^2 \\ &< \lim_{k\to+\infty} P^{x_k}(0)^2 - \lim_{k\to+\infty} P^{x_k}(0)^2 = 0. \end{aligned}$$

Therefore, $0 < \lim_{k\to+\infty} g_6^k(a_k) \leq \lim_{k\to+\infty} g_6^k(\bar{a}) < 0$, which is a contradiction. This result implies that $\lim_{k\to+\infty} P^{x_k}(0) = \|x_k - T(x_k))\|_G^2 = 0$ and, therefore, that every limit point of the sequence $\{x_k\}$ is a fixed point solution. □

Finally, Lemma 6 together with the nonexpansiveness of T imply assumptions A1 and A2.

Appendix B - Optimal convergence rates for affine problems

Using Theorem 6, we will show that under appropriate conditions, when applied to affine maps, the six potential optimization schemes have a better rate of convergence than function iteration. In this discussion, we let a_k^j and $A_k^j(x^*)$ with $j = I, ..., V$ denote the step size and the quantity $A_k(x^*)$ for Scheme j at the kth iteration.

Throughout this discussion, we assume $I - T = M$ for some matrix M and let $T(x_k) - x_k \equiv d_k = -Mx_k$ so that $x_k(a) - T(x_k(a) = Mx_k(a) = Mx_k + aM(T(x_k) - x_k) = -d_k + aMd_k$.

Notice that since $I - T = M$, the fixed points of the map T are any solution to the homogeneous linear system $Mx = 0$. In particular, $x^* = 0$ is a fixed point, which we will use in some of our bounds (e.g., the use of Proposition 1). It is easy to extend this analysis when $x - T(x) = Mx - c$ for some given vector c so that fixed points are solutions to the linear system $Mx = c$.

- **T is firmly contractive and $\beta < \frac{1-A}{2}$ for Scheme I**. Let $\bar{G} = M^tM + \beta I$ in the definition of $A_k(x^*)$ and let $G = I$ in the potential function. Then $g_1^{x_k}(a) = \|x_k(a)-T(x_k(a)))\|^2+a^2\beta\|x_k-T(x_k))\|^2 = \|d_k-aMd_k\|^2+a^2\beta\|d_k\|^2$, which upon expanding and setting the derivative with respect a to zero gives, $a_k^I = \frac{d_k^tMd_k}{\|Md_k\|^2+\beta\|d_k\|^2}$. If T is a firmly contractive map, that is,

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

then the fact that $x - T(x) = Mx$ implies, by expanding $\|T(x) - T(y)\|^2$, that

$$(x - y)^tM(x - y) \geq \frac{1 - A}{2}\|x - y\|^2 + ||Mx - My\|^2.$$

Since Proposition 1 implies that $\|Md_k\| \leq (1 + \sqrt{A})\|d_k\|$, then if $\beta < \frac{1-A}{2}$,

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

In this case, $\frac{d_k^t\bar{G}(x^*-x_k)}{\|d_k\|_{\bar{G}}^2} = \frac{d_k^tMd_k+\beta(x_k-x^*)^tM(x_k-x^*)}{\|Md_k\|^2+\beta\|d_k\|^2}$. Since T is firmly contractive, the second term in the numerator is nonnegative and so $\frac{d_k^t\bar{G}(x^*-x_k)}{\|d_k\|_{\bar{G}}^2} \geq a_k^I$ and, therefore, relation (7) implies that

$$(a_k^I - 1)(A_k^I(x^*) - 1) \geq (a_k^I - 1)(a_k^I - 1) \geq q^2.$$

Therefore, the condition of Theorem 6 holds and so optimizing the potential $g_1^x(a)$ has a better rate of convergence than function iteration.

- **T is contractive for Scheme II**. Observe that a choice of $\bar{G} = M^tM$ and $G = I$ in $g_2^{x_k}(a)$ implies from relation (7) that $A_k(x^*) = \frac{-2d_k^tM^tM(x_k-x^*)}{\|d_k\|_{\bar{G}}^2} - a_k^{II} = \frac{-2d_k^tM^tM(x_k-x^*)}{\|Md_k\|^2} - a_k^{II}$. But since $a_k^{II} = \frac{d_k^tM^td_k}{\|Md_k\|^2} = \frac{d_k^tM^tM(x^*-x_k)}{\|Md_k\|^2}$, $A_k(x^*) = a_k^{II}$ and so $(a_k^{II} - 1)(A_k^{II}(x^*) - 1) = (a_k^{II} - 1)^2 \geq 0$. Therefore, Scheme II always has a guaranteed rate of convergence at least a good as function iteration. Furthermore, if the matrix $M^tM - M$ is strongly positive definite (i.e. for all x, $x^t(M^tM - M)x \geq C\|x\|^2$, for some $C > 0$) then since $K_1\|x\|^2_{\bar{G}} \leq \|x\|^2 \leq K_2\|x\|^2_{\bar{G}}$ for some positive constants K_1 and K_2 and $\bar{G} = M^tM$, $1 - \frac{x^tMx}{\|x\|^2_{\bar{G}}} \geq C\frac{\|x\|^2}{\|x\|^2_{\bar{G}}} \geq K_1C \equiv q$. Therefore, $(a_k^{II} - 1)(A_k^{II}(x^*) - 1) = (\frac{d_k^tM^td_k}{\|Md_k\|^2} - 1)^2 \geq q^2 > 0$, satisfying the condition of Theorem 6. In this case, $a_k^{II} = A_k^{II}(x^*) \leq 1 - K_1C$ are bounded away from 1 from below. If $M - M^M$ is strongly positive definite (i.e. for all x, $x^t(M - M^tM)x \geq C\|x\|^2$, for some $C > 0$), we obtain a similar result with $a_k^{II} = A_k^{II}(x^*) \geq 1 + K_1C > 0$ bounded away from 1 from above and so $(a_k^{II} - 1)(A_k^{II}(x^*) - 1) \geq q^2 > 0$ with $q \equiv K_1C$.

Note that for our example $T(y, z) = b(z, y)$, when $b < 1$, the matrix $M^tM - M = b\begin{bmatrix} b & -1 \\ 1 & b \end{bmatrix}$ is strongly positive definite (with $C = b^2$) and so the result we have just established applies. In a simple single dimensional case, when $T(x) = \frac{1}{2}x$, the matrix $M^tM - M = -\frac{1}{4}$ is strongly negative definite. In this instance $M = \frac{1}{2}$, the optimal step size is 2, and the potential optimization procedure finds a fixed point in one step.

- ***T* is contractive for Scheme III**. The argument is very similar to that for Scheme I. For $G = I$, as in Scheme I, $a_k^{III} = \frac{d_k^tMd_k}{\|Md_k\|^2 - \beta\|d_k\|^2}$, if $\|Md_k\|^2 > \beta\|d_k\|^2$ and $\frac{d_k^tMd_k}{\|Md_k\|^2 - \beta\|d_k\|^2} < c_1$. Otherwise $a_k^{III} = c_1$.

  Observe that $1 - a_k^{III} \geq 1 - c_1 \equiv q > 0$. If $\bar{G} = M^tM - \beta I$ then $\frac{d_k^t\bar{G}(x^* - x_k)}{\|d_k\|^2_{\bar{G}}} = \frac{d_k^tMd_k - \beta(x_k - x^*)^tM(x_k - x^*)}{\|Md_k\|^2 - \beta\|d_k\|^2}$. Since T is firmly contractive, the second term in the numerator is nonnegative and so relation (7) implies that $A_k^{III}(x^*) \leq a_k^{III}$ and therefore $(1 - a_k^{III})(1 - A_k^{III}(x^*)) \geq q^2$, satisfying the condition of Theorem 6.

- ***T* is firmly contractive for Scheme IV**. If $G = I$, then $a_k^{IV} = \frac{\beta\|d_k\|^2 + 2d_k^tMd_k}{2\beta\|d_k\|^2 + 2\|Md_k\|^2} \geq \frac{\beta\|d_k\|^2 + 2d_k^tMd_k}{2\|Md_k\|^2}$. Arguing as in the analysis of Scheme I shows that

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

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

  For a choice of $\bar{G} = M^tM$, $\frac{d_k^t\bar{G}(x^* - x_k)}{\|d_k\|^2_{\bar{G}}} = \frac{d_k^tMd_k}{\|Md_k\|^2}$. If T is firmly contractive, then relation (7), as in the discussion for Scheme I, implies that

$$\begin{aligned} A_k^{IV}(x^*) - 1 &= \frac{2d_k^tMd_k - \|Md_k\|^2}{\|Md_k\|^2} - a_k^{IV} \\ &\geq \frac{d_k^tMd_k + \frac{1-A}{2}\|d_k\|^2}{\|Md_k\|^2} - a_k^{IV} \geq \frac{(\frac{1-A-\beta}{2})\|d_k\|^2}{\|Md_k\|^2} \geq q, \end{aligned}$$

for all $\beta < 1 - A$. Therefore, a choice of $\beta < 1 - A$ guarantees that $(a_k^{IV} - 1)(A_k^{IV}(x^*) - 1) \geq q^2$, satisfying the condition of Theorem 6.

- ***T* is contractive and *M* is symmetric for Scheme V (that is, for the steepest descent method)** Since $M = M^t$, choices of $\bar{G} = M$ and $G = I$ imply that $a_k^V = \frac{d_k^t M(x^* - x_k)}{\|d_k\|_M^2} = \frac{\|d_k\|^2}{\|d_k\|_M^2} = A_k^V(x^*)$. This scheme always at least as good a rate of convergence as function iteration since $(a_k^V - 1)(A_k^V(x^*) - 1) = (a_k^V - 1)^2 \geq 0$. If the matrix $M - I$ is strongly positive definite (i.e. for all $x, x^t(M - I)x \geq C\|x\|^2$, for some $C > 0$) then for an appropriate constant K_1, as in the analysis of Scheme II with $\bar{G} = M$, $(a_k^V - 1)(A_k^V(x^*) - 1) = (\frac{d_k^t d_k}{\|d_k\|_M^2} - 1)^2 \geq (K_1 C)^2 > 0$. If $I - M$ is strongly positive definite, then we obtain a similar result and bound $(a_k^{II} - 1)(A_k^{II}(x^*) - 1) \geq (K_1 C)^2 > 0$.