Proc. Indian Acad. Sci. (Math. Sci.), Vol. 106, No. 3, August 1996, pp. 289–300.

Distributed computation of fixed points of ∞-nonexpansive maps

VIVEK S BORKAR

Department of Computer Science and Automation, Indian Institute of Science, Bangalore 560 012, India

MS received 7 August 1995; revised 3 May 1996

Abstract. The distributed implementation of an algorithm for computing fixed points of an ∞-nonexpansive map is shown to converge to the set of fixed points under very general conditions.

Keywords. Distributed algorithm; fixed point computation; ∞-nonexpansive map; tapering stepsize; controlled Markov chains.

1. Introduction

Many problems in optimization can be cast as problems of finding a fixed point of a map $F:R^d \to R^d$ which is nonexpansive with respect to the ∞-norm, or a weighted version thereof. Examples are algorithms for solving dynamic programming equations for shortest path problems and Markov decision processes, certain network flow problems, etc. A standard approach then is to use the iteration

$$x_{n+1} = F(x_n)$$

or its 'relaxed' version

$$x_{n+1} = (1-a)x_n + aF(x_n), \quad a \in (0,1). \tag{1}$$

A comprehensive account of synchronous and asynchronous implementations of these algorithms and their applications to optimization and numerical analysis appears in [1], along with an extensive bibliography. A continuous time analog is studied in [4].

This work considers a distributed implementation of a variant of (1). We replace 'a' by a tapering stepsize $\{a(n)\} \subset (0,1)$ as in stochastic approximation theory, satisfying for some $r \in (0,1)$,

$$1 \geqslant a(n+1)/a(n) \to 1, \quad \sum_n a(n) = \infty, \quad \sum_n a(n)^{1+q} < \infty \tag{2}$$

for $q \geqslant r$. An example is $a(n) = (n+2)^{-1}$, $r = $ (say) 0·5. Our model of distributed computation is as follows: We postulate a set-valued random process which selects at each time a subset of $\{1,\ldots,d\}$ indicating the indices of the components which are to be updated. The update uses delayed information regarding other components with the delays being required to satisfy a mild conditional moment bound. We obtain a very general convergence theorem under these conditions (Theorem 1.1 below) which does not require F to be a contraction or a pseudocontraction as in the earlier literature.

Tapering stepsize as a means for suppressing the effects of delays was proposed in [3] in the case of algorithms whose continuous limits are differential equations admitting strict Liapunov functions. Equation (1) need not satisfy this condition. Furthermore, the model of distributed computing is more elaborate than in the above work.

The paper is organized as follows. The remainder of this section formulates the algorithm and states the main result. The next section studies an associated o.d.e. which this algorithm tracks in the limit. The third section proves the main result using this 'o.d.e. limit'. The last section gives some examples if nonexpansive maps arising in numerical analysis and optimization. A forthcoming companion paper studies the asynchronous version of this algorithm (i.e., one without a 'global clock').

Introduce the norms

$$\|x\|_p = \left(\frac{1}{d}\sum_{i=1}^{d} |x_i|^p\right)^{1/p}, \quad 1 \leqslant p < \infty,$$

$$\|x\|_\infty = \max_i |x_i|.$$

for $x = [x_1, \ldots, x_d]$. Let $F: R^d \to R^d$ be ∞-nonexpansive, i.e.,

$$\|F(x) - F(y)\|_\infty \leqslant \|x - y\|_\infty, \quad x, y \in R^d.$$

In particular, it is Lipschitz. Let $G = \{x | F(x) = x\}$ denote the (closed) set of fixed points of F, assumed nonempty.

Let $I = \{1, 2, \ldots, d\}$ and S a collection of nonempty subsets of I that cover I. Let $\{Y_n\}$ be an S-valued process and for each n, $\tau_{ij}(n)$, $i \neq j \in I$, random variables (delays) taking values $\{0, 1, \ldots, n\}$. We set $\tau_{ii}(n) = 0\ \forall i, n$. The distributed version of (1) we consider is as follows: Given $X(0)$, compute $X(n) = [X_1(n), \ldots, X_d(n)]$ iteratively by

$$\begin{aligned} X_i(n+1) = X_i(n) + a(n)[F_i(X_1(n - \tau_{i1}(n)), \ldots, X_d(n - \tau_{id}(n))) \\ - X_i(n)] I\{i \in Y_n\} \end{aligned} \tag{3}$$

for $i \in I$, $n \geqslant 0$. Let $\mathscr{F}_n = \sigma(X(m), Y(m), m \leqslant n, \tau_{ij}(m), m < n, i, j \in I)$ and $\mathscr{G}_n = \sigma(X(m), Y(m), \tau_{ij}(m), m \leqslant n, i, j \in I)$, $n \geqslant 0$. We shall be making the following key assumptions:

(A1) There exists a $\delta > 0$ such that the following holds. For any $A, B \in S$, the quantity

$$P(Y_{n+1} = B / Y_n = A, \mathscr{G}_n) \tag{4}$$

is either always zero, a.s., or always exceeds δ, a.s. That is, having picked A at time n, picking B at the next instant is either improbable or probable with a minimum conditional probability of δ, regardless of n and the 'history' $\mathscr{G}_n$.

Furthermore, if we draw a directed graph with node set S and an edge from A to B when (4) exceeds δ a.s., then this graph is irreducible, i.e., there is a directed path from any $A \in S$ to any $B \in S$.

(A2) There exist $b \geqslant r/(1-r)$ and $C > 0$ such that

$$E[(\tau_{ij}(n))^b / \mathscr{F}_n] \leqslant C \quad \text{a.s.} \quad \forall i, j, n.$$

(A3) If $\bar{n}$ = the integer part of $a(n)^{r-1}$, then $\bar{n}$ is $o(n)$ and moreover, $\limsup_{n\to\infty} a(n - \bar{n})/a(n) < \infty$. (Note that this condition is satisfied by our example $a(n) = (n+2)^{-1}$ with $r = 0{\cdot}5$.)

The condition $\tau_{ii}(n) = 0$ implies that for updating the ith component, its most recent value is immediately available. The idea is that each component is computed by a specific processor, and the different processors communicate with each other in conformity with (A2). We make the following immediate observation for later use.

Let $z \in G$. If $C_n = \max_{m \leq n} \|X(n) - z\|_\infty$, then by (3) and ∞-nonexpansivity $F, \{C_n\}$ is a nonincreasing sequence. Thus

$$\sup_n \|X(n) - z\|_\infty \leq \|X(0) - z\|_\infty.$$

Our main result is the following:

Theorem 1.1. $X(n) \to G$ *a.s.*

The proof will use the fact that (3) asymptotically tracks an o.d.e. The next section studies this o.d.e.

2. An associated o.d.e.

We start with some notation. For $A \in S$, define $F^A(\cdot) = [F_1^A(\cdot) \ldots, F_d^A(\cdot)]$: $R^d \to R^d$ by

$$F_j^A(x) = F_j(x) I\{j \in A\} + x_j I\{j \notin A\}$$

for $x = [x_1, \ldots, x_d]$. For any Polish space X, let $\mathscr{P}(X)$ denote the Polish space of probability measures on X with the Prohorov topology. In particular, $\mathscr{P}(S)$ is the space of probability vectors on S. For $\mu \in \mathscr{P}(S)$, define F^μ: $R^d \to R^d$ by

$$F^\mu(\cdot) = \sum \mu(A) F^A(\cdot).$$

Clearly, F^A, F^μ are ∞-nonexpansive. Given $\alpha > 0$, say that $\mu \in \mathscr{P}(S)$ is α-thick if $\min_A \mu(A) \geq \alpha$ and thick if it is α-thick for some $\alpha > 0$. The following is easily proved.

Lemma 2.1. *G is precisely the set of fixed points of F^μ for thick μ and the intersection of the sets of fixed points of F^A, $A \in S$ (resp., F^μ, $\mu \in \mathscr{P}(S)$).*

Let U denote the space of $\mathscr{P}(S)$-valued trajectories $\bar{\mu} = \{\mu_t, t \geq 0\}$ with the coarsest topology that renders continuous the maps $\bar{\mu} \to \int_0^T f(t) \mu_t(A) \mathrm{d}t$ for $T \geq 0$, $A \in S$, $f \in L_2[0, T]$. U then is compact metrizable. Let us say that $\bar{\mu} \in U$ is α-thick for a given $\alpha > 0$ if μ_t is for a.e. t, where the 'a.e.' may be dropped by taking an appropriate version. Say that it is thick if it is α-thick for some $\alpha > 0$.

Lemma 2.2. *Let $\bar{\mu}^n \to \bar{\mu}^\infty$ in U and $\bar{\mu}^n$ is α-thick, $\alpha > 0$, for $n = 1, 2, \ldots$. Then $\bar{\mu}^\infty$ is α-thick.*

Proof. For any $A \in S$, $t > s \geq 0$, $n \geq 1$,

$$\int_s^t \mu_y^n(A) \mathrm{d}y \geq \alpha(t - s).$$

The inequality is preserved in the limit $n \to \infty$. The rest is easy. □

Given $\bar{\mu} \in U$, consider the o.d.e.

$$\dot{x}(t) = F^{\mu_t}(x(t)) - x(t), \; x(0) = x. \tag{5}$$

Lemma 2.3. *The map $(\bar{\mu}, x) \in U \times R^d \to x(\cdot) \in C([0, \infty); R^d)$ defined by* (5) *is continuous.*

Proof. Let $(\bar{\mu}^n, x_n) \to (\bar{\mu}^\infty, x_\infty)$. For $n \geq 1$, let $x^n(\cdot)$ satisfy

$$\dot{x}^n(t) = F^{\mu_t^n}(x^n(t)) - x^n(t), \; x^n(0) = x_n.$$

Using the Gronwall lemma and Arzela-Ascoli theorem, one verifies that $\{x^n(\cdot)\}$ is relatively compact in $C([0,\infty); R^d)$. By dropping to a subsequence if necessary, suppose that $x^n(\cdot)\to x^\infty(\cdot)$. Then $x^\infty(0)=x_\infty$. Also, for $t\geqslant 0$, $n\geqslant 1$,

$$\begin{aligned} x^n(t) = x_n &+ \int_0^t (F^{\mu^n_s}(x^n(s)) - F^{\mu^n_s}(x^\infty(s)))\mathrm{d}s \\ &+ \int_0^t (F^{\mu^n_s}(x^\infty(s)) - F^{\mu^\infty_s}(x^\infty(s)))\mathrm{d}s \\ &+ \int_0^t (F^{\mu^\infty_s}(x^\infty(s)) - x^\infty(s))\mathrm{d}s + \int_0^t (x^\infty(s) - x^n(s))\mathrm{d}s. \end{aligned} \tag{6}$$

The first and the fourth integral go to zero as $n\to\infty$ because $x^n(\cdot)\to x^\infty(\cdot)$. (We use the ∞-nonexpansivity of F^μ in the former case.) So does the second integral in view of our topology on U. Thus $x^\infty(\cdot)$ satisfies

$$\dot{x}^\infty(t) = F^{\mu^\infty_t}(x^\infty(t)) - x^\infty(t),\ x^\infty(0) = x_\infty.$$

The claim follows. □

The main result of this section is the following:

Theorem 2.1. *Given a thick $\bar{\mu}\in U$, the solution $x(\cdot)$ of (5) converges to a point in G which may depend on the initial condition. Furthermore, $t\to\|x(t)-x^*\|_\infty$ is nonincreasing for any $x^*\in G$.*

The proof will closely follow that of Theorem 3.1 [4], except for the additional complications caused by the fact that (5) is nonautonomous. We split the proof into several lemmas. Let $x^*\in G$ and $\bar{\mu}$ be α-thick for $\alpha>0$.

Lemma 2.4. *The map $t\to\|x(t)-x^*\|_\infty$ is nonincreasing and hence converges to some $a>0$.*

Proof. A straightforward computation shows that for $p\in(1,\infty)$

$$\frac{\mathrm{d}}{\mathrm{d}t}\|x(t)-x^*\|_p = -\|x(t)-x^*\|_p + \|x(t)-x^*\|_p^{1-p} g(\mathrm{t}),$$

where

$$\begin{aligned} g(t) &= \frac{1}{d}\sum_{i=1}^{d} |x_i(t)-x_i^*|^{p-1}\, sgn(x_i(t)-x_i^*)(F_i^{\mu_t}(x(t)) - F_i^{\mu_t}(x^*)) \\ &\leqslant \|x(t)-x^*\|_p^{p-1}\|F^{\mu_t}(x(t)) - F^{\mu_t}(x^*)\|_p. \end{aligned} \tag{7}$$

Thus for $t>s$,

$$\begin{aligned} \|x(t)-x^*\|_p \leqslant \|x(s)-x^*\|_p + \int_s^t (&-\|x(y)-x^*\|_p \\ &+ \|F^{\mu_y}(x(y)) - F^{\mu_y}(x^*)\|_p)\mathrm{d}y. \end{aligned}$$

Let $p\to\infty$ to obtain

$$\begin{aligned} \|x(t)-x^*\|_\infty \leqslant \|x(s)-x^*\|_\infty + \int_s^t (&-\|x(y)-x^*\|_\infty \\ &+ \|F^{\mu_y}(x(y)) - F^{\mu_y}(x^*)\|_\infty)\mathrm{d}y. \end{aligned}$$

The claim follows in view of the ∞-nonexpansivity of F^{μ}. □

If $a = 0$, we are done. Suppose $a > 0$. Define $B_a = \{x \in R^d | \|x - x^*\|_\infty = a\}$. Introduce the terminology: An m-face for $m \leq d$ is a set of the type

$$\{x = [x_1, \ldots, x_d] | x_{i_k} \in [a_{i_k}, b_{i_k}], k \leq m, x_{i_k} = c_{i_k}, k > m\},$$

where $\{i_1, \ldots, i_d\}$ is a permutation of $\{1, \ldots, d\}$ and $b_k > a_k$, c_k are scalars. Then B_a is the union of $(d-1)$-faces of the type $\{x | x_i - x_i^* = a \text{ (or } -a) \text{ and } |x_j - x_j^*| < a \text{ for } j \neq i\}$. For any $(d-1)$-face of this type (say, H), define $G_H = \{x \in H | F(x) \in H\}$. Then G_H is closed, possibly empty. Since $\|x(t) - x^*\|_\infty \to a$, $x(t) \to B_a$. Now the trajectories $x(t + \cdot)$, $t \geq 0$, form a relatively compact set in $C([0, \infty); R^d)$. Thus any limit point $\tilde{x}(\cdot)$ thereof as $t \to \infty$ must lie in B_a. By Lemmas 2.2 and 2.3, $\tilde{x}(\cdot)$ satisfies

$$\dot{\tilde{x}}(t) = F^{\bar{\mu}_t}(\tilde{x}(t)) - \tilde{x}(t)$$

for some α-thick $\bar{\mu} \in U$. Let $\{\tilde{x}(\cdot)\} = \{\tilde{x}(t) | t \in R^+\}$.

Lemma 2.5. $\{\tilde{x}(\cdot)\} \cap H \subset G_H$.

Proof. If both sets are empty, there is nothing to prove. Suppose $\{\tilde{x}(\cdot)\} \cap H \neq \phi$. For simplicity, let $H = \{x | x_1 - x_1^* = a, |x_i - x_i^*| \leq a, i > 1\}$. Suppose $\{\tilde{x}(t) | t \in [0, \Delta]\} \subset H$. Then $\tilde{x}_1(t) = x_1^* + a$, $t \in [0, \Delta]$, leading to $0 = \dot{\tilde{x}}_1(t) = F_1^{\bar{\mu}_t}(\tilde{x}(t)) - \tilde{x}_1(t)$. Thus $F_1^{\bar{\mu}_t}(\tilde{x}(t)) = \tilde{x}_1(t) = x_1^* + a = F_1(\tilde{x}(t))$ in view of α-thickness of $\bar{\mu}$, for $t \in [0, \Delta]$. Since F is ∞-nonexpansive and x^* is a fixed point of it, we also have

$$\|F(\tilde{x}(t)) - x^*\|_\infty \leq \|\tilde{x}(t) - x^*\|_\infty = a.$$

Thus we must have $F(\tilde{x}(t)) \in H$, $t \in [0, \Delta]$, implying $\tilde{x}(t) \in G_H$. It follows that all connected segments of $\{\tilde{x}(\cdot)\} \cap H$ that contain more than one point must be in G_H. On the other hand, those containing a single point must clearly be in the relative boundary ∂H of H, which is the union of its $(d-2)$-faces. Let $x \in \{\tilde{x}(\cdot)\} \cap \partial H$. It suffices to show that $F(x) \in \partial H$. If not, $F(x) - x$ and therefore $F^{\mu}(x) - x$ for any thick μ would be transversal to ∂H at x. This is not possible because $\tilde{x}(\cdot)$ is a differentiable trajectory confined to B_a and cannot make 'sharp turns' around corners. This completes the proof. □

Consider a fixed $(d-1)$-face H of B_a for the time being. Let $H = \{x | x_1 = x_1^* + a_1, |x_j - x_j^*| \leq a, j > 1\}$ for simplicity.

Lemma 2.6. *If $G_H \neq \phi$, the map $F: G_H \to H$ can be extended to an ∞-nonexpansive map $\tilde{F}: H \to H$ which has a fixed point $\tilde{x} \in H$.*

Proof. The second claim follows from the first by the Brouwer fixed point theorem. Fix $1 < i \leq d$ and define

$$g_i(x) = \inf_{y \in G_H} (F_i(y) + \|x - y\|_\infty), \; x \in H.$$

Then $g_i(x) \leq F_i(x)$, $x \in G_H$. For $x, y \in G_H$, ∞-nonexpansivity of F leads to

$$F_i(y) + \|x - y\|_\infty \geq F_i(x).$$

Thus $g_i(x) \geqslant F_i(x)$, implying $g_i = F_i$ on G_H. Now for $x, z \in H$,

$$g_i(x) \leqslant \inf_{y \in G_H} (F_i(y) + \|y - z\|_\infty + \|z - x\|_\infty)$$

$$\leqslant g_i(z) + \|z - x\|_\infty. \tag{8}$$

Similarly, $g_i(z) \leqslant g_i(x) + \|z - x\|_\infty$. Hence

$$|g_i(x) - g_i(z)| \leqslant \|z - x\|_\infty.$$

Let $\tilde{F}_i(x) = \max(x_i^* - a, \min(g_i(x), x_i^* + a))$. Then

$$|\tilde{F}_i(x) - \tilde{F}_i(z)| \leqslant \|x - z\|_\infty.$$

Let $\tilde{F}_1(x) = x_1^* + a$, $x \in H$. Then $\tilde{F}(\cdot) = [\tilde{F}_1(\cdot), \ldots, \tilde{F}_d(\cdot)]$ is the desired map. □

Proof of Theorem 2.1. The same argument as in Lemma 2.6 can be used to extend $\tilde{F}$ to an ∞-nonexpansive map $\bar{F}: R^d \to R^d$ that restricts to $\tilde{F}$ on H and to F on $\cup_A G_A$. Define $\bar{F}^A$, $\bar{F}^\mu$ in analogy with F^A, F^μ using $\bar{F}$ in place of F. Then $\tilde{x}(\cdot)$ satisfies

$$\dot{\tilde{x}}(t) = \bar{F}^{\bar{\mu}_t}(\tilde{x}(t)) - \tilde{x}(t).$$

We conclude that $t \to \|\tilde{x}(t) - \tilde{x}\|_\infty$ is nonincreasing in t and hence decreases to $b \geqslant 0$. If $b = 0$, we are done. If not, let $B_b = \{x \mid \|x - \tilde{x}\|_\infty = b\}$. Then $\tilde{x}(t) \to B_b$. Also, it is clear that no $(d-1)$-face H of B_b is coplanar with H. This argument can now be repeated for each $(d-1)$-face H of B_a that intersects $\{\tilde{x}(\cdot)\}$, leading to possibly more $\|.\|_\infty$-spheres $B_c, B_d, \ldots$ defined analogously to B_a such that $\tilde{x}(t) \to B_a \cap B_b \cap B_c \cap \ldots$. The above remarks also imply that this intersection is a union of m-faces with m at most $(d-2)$. Now consider a limit point $x'(\cdot)$ of $\tilde{x}(t + .)$, $t \geqslant 0$, in $C([0, \infty); R^d)$ as $t \to \infty$. Repeat the above argument to conclude that $x'(t)$ converges to a union of m-faces with m at most $(d-3)$. Iterating this argument at most d times, we are left with a union of finitely many points to one of which $\tilde{x}(\cdot), x'(\cdot) \ldots$ and hence $x(\cdot)$ must converge and which then must be a fixed point of F^μ for some thick μ, hence of F. This completes the proof. □

COROLLARY 2.1

Given $\varepsilon, b > 0$, *there exist* $T = T(\varepsilon, b) > 0$, $\eta = \eta(\varepsilon, b) > 0$ *such that for any solution* $x(\cdot)$ *of* (5) *satisfying* $\|X(0)\|_\infty \leqslant K < \infty$ *and*

(i) $\bar{\mu}$ *is b-thick,*
(ii) $\{x(t) \mid t \in [0, T]\} \cap G^\varepsilon = \phi$ *where* $G^\varepsilon = \{x \mid \inf_{y \in G} \|x - y\| < \varepsilon\}$, *we have*

$$\inf_{y \in G} \|x(t) - y\|_\infty \leqslant \inf_{y \in G} \|x(0) - y\|_\infty - \eta \quad \text{for} \quad t \geqslant T.$$

Proof. Suppose that the claim is false. Then there exist $x^n \in R^d$, $\bar{\mu}^n \in U$ such that for some $b > 0$, $\{\bar{\mu}^n\}$ are b-thick and $x^n(\cdot)$ satisfy (i) $\dot{x}^n(t) = F^{\mu_t^n}(x^n(t)) - x^n(t)$, $x^n(0) = x^n$, $\|x^n\|_\infty \leqslant K$, (ii) $x^n(t) \notin G^\varepsilon$ for $t \in [0, n]$ and

$$\inf_{y \in G} \|x^n - y\|_\infty \geqslant \sup_{t \in [0, n]} \inf_{y \in G} \|x^n(t) - y\|_\infty \geqslant \inf_{y \in G} \|x^n - y\|_\infty - 1/n,$$

for $n \geqslant 1$. By dropping to a subsequence if necessary, we may then suppose that

$x^n \to x^\infty \notin G^\varepsilon$, $\bar{\mu}^n \to \bar{\mu}^\infty$ in U and $x^n(\cdot) \to x^\infty(\cdot)$ in $C([0, \infty); R^d)$. By Lemmas 2.2 and 2.3, $\bar{\mu}^\infty$ is b-thick and $x^\infty(\cdot)$ satisfies.

$$\dot{x}^\infty(t) = F^{\mu_t^\infty}(x^\infty(t)) - x^\infty(t),\ x^\infty(0) = x^\infty.$$

Also, $\inf_{y \in G} \|x^\infty - y\|_\infty = \inf_{y \in G} \|x^\infty(t) - y\|_\infty$, $t \geq 0$. This contradicts Theorem 2.1, establishing the claim. □

For $\alpha, \eta, T > 0$, call a trajectory $y(\cdot): R^+ \to R^d$ an (α, η, T)-perturbation of (5) if there exist $0 = T_0 < T_1 < T_2 < \cdots$ in $[0, \infty)$ with $T_{j+1} - T_j \geq T$, such that for some $x^j(\cdot)$ satisfying (5) for some α-thick $\bar{\mu}^j$ in place of $\bar{\mu}$, we have

$$\sup_{t \in [T_j, T_{j+1}]} \|y(t) - x^j(t)\|_\infty < \eta, j \geq 0.$$

COROLLARY 2.2

For any $\alpha, \varepsilon > 0$, $T > 0$ as in Corollary 2.1 and $\gamma > 0$ sufficiently small, any (α, γ, T)-perturbation $y(\cdot)$ of (5) converges to G^ε.

Proof. In view of Corollary 2.1, this is a straightforward adaptation of Theorem 1, p. 339 of [5]. □

3. Proof of Theorem 1.1

As a first step towards establishing that $\{X(n)\}$ tracks (5) asymptotically, we analyze the S-valued process $\{Y_n\}$. In particular, we shall show that it may be viewed as a controlled Markov chain.

For $A \in S$, let $D_A = \{B \in S |$ (4) exceeds δ, a.s.$\}$ and $V_A = \{u \in \mathscr{P}(D_A) | u(B) \geq \delta\, \forall B \in D_A\}$. Let $V = \Pi_A V_A$. Define $p: S \times S \times V \to [0, 1]$ by

$$p(A, B, \mu) = \mu_A(B),$$

where μ_A is the Ath component of μ. Define V-valued random variables $\{Z^n\}$ as follows. The Ath component of Z^n, denoted by Z^n_A is given by

$$Z^n_A(B) = P(Y_{n+1} = B/\mathscr{G}_n) I\{Y_n = A\} + \Psi_A I\{Y_n \neq A\},$$

where Ψ_A are fixed elements of V_A, $A \in S$. Then (4) equals $p(A, B, Z^n)$ and $\{Y_n)$ may be viewed as an S-valued controlled Markov chain with action space V and transition probability function p. It should be kept in mind, however, that this is purely a technical convenience and it is in no way implied that $\{Z^n\}$ is actually a control process. In particular, this allows us to conceive of a 'stationary control policy π' associated with a map $\pi: S \to V$ wherein $Z^n = \pi(Y_n)$, $n \geq 0$. The latter part of (A1) implies that $\{Y_n\}$ will be an ergodic Markov chain under any stationary policy π with a corresponding unique stationary distribution $v_\pi \in \mathscr{P}(S)$.

Let $t_0 = 0$, $t_n = \Sigma_{m=1}^n a(m)$, $n \geq 1$. Define $y(\cdot): [0, \infty) \to S$ by

$$y(t) = Y_n, t_n \leq t < t_{n+1}, n \geq 0.$$

Define $\bar{\mu} \in U$ by $\mu_t(A) = I\{y(t) = A\}$, $t \geq 0$ and $\bar{\mu}^s \in U$, $s \geq 0$, by $\mu^s_t = \mu_{s+t}$, $t \geq 0$.

Lemma 3.1. *There exists an $\alpha > 0$ such that any limit point of $\{\bar{\mu}^s\}$ as $s \to \infty$ is α-thick, a.s.*

Proof. Let $A \in S$. Then

$$M_n = \sum_{m=1}^{n} a(m)[I\{Y_m = A\} - \sum_B p(B, A, Z^{m-1}) I\{Y_{m-1} = B\}]$$

is a zero mean bounded increment martingale with respect to $\{\mathscr{G}_n\}$, whose quadratic variation process converges a.s. in view of (2). By Proposition VII-2-3(c), pp. 149–150, [6], $\{M_n\}$ converges a.s. For $n \geq 0$, let $\bar{n}(s) = \min\{m > n \mid \Sigma_{j=n+1}^{m} a(j) \geq s\}$, $s > 0$. Then $\lim_{n\to\infty}(M_{\bar{n}(s)} - M_n) = 0$ a.s. and $\Sigma_{m=n}^{\bar{n}(s)} a(m) \geq s$ together imply

$$\frac{\sum_{m=n}^{\bar{n}(s)} a(m) I\{Y_m = A\}}{\sum_{m=n}^{\bar{n}(s)} a(m)} - \frac{\sum_{m=n}^{\bar{n}(s)} a(m) \sum_B p(B, A, Z^{m-1}) I\{Y_{m-1} = B\}}{\sum_{m=n}^{\bar{n}(s)} a(m)} \to 0 \tag{9}$$

a.s. Define $\Phi_{n,s} \in \mathscr{P}(S \times V)$ by

$$\Phi_{n,s}(C \times J) = \frac{\sum_{m=n}^{\bar{n}(s)} a(m) I\{Y_m \in C, Z^m \in J\}}{\sum_{m=n}^{\bar{n}(s)} a(m)}$$

for $C \subset S$, $J \subset V$ Borel. Then from (9), it follows that a.s., any limit point Φ of $\Phi_{n,s}$ in $\mathscr{P}(S \times V)$ as $n \to \infty$ must satisfy

$$\Phi(\{A\} \times V) = \int p(.,A,.) \mathrm{d}\Phi, \; A \in S.$$

Thus Φ must be of the form

$$\Phi(\{A\} \times J) = v(A) \varphi_A(J), \; A \in S, J \subset V \text{ Borel},$$

where $A \to \varphi_A : S \to \mathscr{P}(V)$ defines a stationary randomized policy and v the corresponding stationary distribution (see e.g. [2], pp. 55–56). By Lemma 1.2, p. 56 and Lemma 2.1, p. 60 of [2], it follows that

$$\min_A v(A) \geq \min_{A,\pi} v_\pi(A) \triangleq \alpha > 0.$$

Thus

$$\liminf_{n\to\infty} \frac{\sum_{m=n}^{\bar{n}(s)} a(m) I\{Y_m = A\}}{\sum_{m=n}^{\bar{n}(s)} a(m)} \geq \alpha \text{ a.s.}$$

From our definition of $\{\bar{\mu}^t\}$, it then follows that

$$\liminf_{t\to\infty} \frac{1}{s} \int_0^s \mathrm{d}y \mu_y^t(A) \geq \alpha \text{ a.s.} \forall A \in S.$$

Fix a sample point outside the zero probability set on which this claim fails for any $A \in S$, $s > 0$ rational. For this sample point, the claim follows easily in view of our topology on U. □

Now rewrite (3) as

$$X(n+1) = X(n) + a(n)(W(n) - X(n))$$

for appropriately defined $W(n) = [W_1(n), \ldots, W_d(n)]$ and set $\hat{W}(n) = E[W(n)/\mathscr{F}_n]$, $n \geq 0$, the conditioning being componentwise. Write $\hat{W}(n) = [\hat{W}_1(n), \ldots, \hat{W}_d(n)]$.

Lemma 3.2. *There exist $K>0$, $N\geqslant 1$ such that for $n\geqslant N$,*

$$\|F^{Y_n}(X(n))-\hat{W}(n)\|_\infty < Ka(n)^r.$$

Proof. W.l.o.g., let $X(0)$ be deterministic. Then $\forall A\subset I$, $z\in G$,

$$\|F^A(X(n))\|_\infty \leqslant \|X(n)-z\|_\infty + \|F(z)\|_\infty V\|z\|_\infty$$
$$\leqslant 3\|z\|_\infty + \|F(z)\|_\infty + \|X(O)\|_\infty \triangleq M<\infty.$$

For each i, $1\leqslant i\leqslant d$, and $c=1-r$,

$$|F^{Y_n}(X(n))-\hat{W}_i(n)| \leqslant E[|F_i^{Y_n}(X(n))-W_i(n)|I\{\tau_{ij}(n)\leqslant a(n)^{-c} \text{ for all } i,j\}/\mathscr{F}_n]$$
$$+E[|F_i^{Y_n}((X(n))-W_i(n)|I\{\tau_{ij}(n)>a(n)^{-c} \text{ for some } i,j\}/\mathscr{F}_n].$$

The second term is bounded by $2MCa(n)^{bc}$ in view of (A2) and the conditional Chebyshev inequality. Let $\bar{n}$ be the integer part of $a(n)^{-c}$. Since $a(n)^{-c}$ is $o(n)$, we may pick n large enough so that $n>\bar{n}$. Then for $m\leqslant\bar{n}$, (A3) leads to

$$\|X(n)-X(n-m)\|_\infty \leqslant 2M\sum_{k=n-\bar{n}}^{n} a(k)\leqslant \bar{K}a(n)^{1-c}$$

for a suitable $\bar{K}>0$. Thus the first term is bounded by $\bar{K}a(n)^r$. Since $b\geqslant r/(1-r)$, the claim follows. □

Let $T>0$. Define $T_0=0, T_n=\min\{t_m|t_m\geqslant T_{n-1}+T\}$, $n\geqslant 1$. Then $T_n=t_{m(n)}$ for a strictly increasing sequence $\{m(n)\}$. Let $I_n=[T_n,T_{n+1}]$, $n\geqslant 0$. Define $\bar{x}^n(t)$, $t\in I_n$, by $\bar{x}^n(T_n)=X(m(n))$ and

$$\bar{x}^n(t_{m(n)+k+1})=\bar{x}^n(t_{m(n)+k})+F^{Y_{m(n)+k}}(\bar{x}^n(t_{m(n)+k}))$$
$$-\bar{x}^n(t_{m(n)+k}))(t_{m(n)+k+1}-t_{m(n)+k}),$$

with linear interpolation on each interval $[t_{m(n)+k},t_{m(n)+k+1}]$. Define $x(t)$, $t\geqslant 0$, by $x(t_n)=X(n)$ with linear interpolation on each interval $[t_n,t_{n+1}]$.

Lemma 3.3.

$$\lim_{n\to\infty}\sup_{t\in I_n}\|x(t)-\bar{x}^n(t)\|_\infty=0 \text{ a.s.}$$

Proof. Let $n\geqslant 1$. For $i\geqslant m(n)$, we have

$$x(t_{i+1})=x(t_i)+a(i)(F^{Y_i}(x(t_i))-x(t_i))+a(i)(\hat{W}(i)-F^{Y_i}(x(t_i)))$$
$$+a(i)(W(i)-\hat{W}(i)).$$

Let $\bar{M}_m=\Sigma_{i=0}^m a(i)(W(i)-\hat{W}(i))$ and $\xi_i=\bar{M}_i-\bar{M}_{m(n)}$ for $i\geqslant m(n)$. Then $\{\bar{M}_m,\mathscr{F}_m\}$ is a zero mean bounded increment vector martingale and the quadratic variation process of each of its component martingales converges a.s. by virtue of (2). Thus by Proposition VII-2-3(c), pp. 149–150 of [6], $\{\bar{M}_m\}$ converges a.s. Fix a sample point for which this convergence holds and let $\delta>0$. Then

$$\sup_{i\geqslant m(n)}\|\xi_i\|_\infty<\delta/2$$

for n sufficiently large.

Let $\hat{x}_{i+1} = x(t_{i+1}) - \xi_i, i \geqslant m(n)$ with $\hat{x}_{m(n)} = X_{m(n)}$ (i.e., $\xi_{m(n)-1} \triangleq 0$). Then for $i \geqslant m(n)$,

$$\hat{x}_{i+1} = \hat{x}_i + a(i)(F^{Y_i}(\hat{x}_i) - \hat{x}_i) + a(i)(F^{Y_i}(\hat{x}_i + \xi_{i-1}) - (\hat{x}_i + \xi_{i-1}) - F^{Y_i}(\hat{x}_i) + \hat{x}_i) + a(i)(\hat{W}(i) - F^{Y_i}(x(t_i))).$$

Also

$$\bar{x}^n(t_{i+1}) = \bar{x}^n(t_i) + a(i)(F^{Y_i}(\bar{x}^n(t_i)) - \bar{x}^n(t_i)).$$

Subtracting and using the preceding lemma, we have, for n sufficiently large,

$$\|\hat{x}_{i+1} - \bar{x}^n(t_{i+1})\|_\infty \leqslant (1 + 2a(i))\|\hat{x}_i - \bar{x}^n(t_i)\|_\infty + 2a(i)\|\xi_{i-1}\|_\infty + Ka(i)^{1+r}.$$

By increasing n if necessary, we may suppose that

$$\sum_{i \geqslant n} a(i)^{1+r} < \delta/2.$$

Then using the inequality $1 + 2a(i) \leqslant \exp(2a(i))$ and iterating, we have for n sufficiently large,

$$\sup_{m(n) \leqslant i \leqslant m(n+1)} \|\hat{x}_i - \bar{x}^n(t_i)\|_\infty \leqslant 2e^{2(T+1)}(K + T + 1)\delta.$$

Also

$$\sup_{m(n) \leqslant i \leqslant m(n+1)} \|\hat{x}_i - x(t_i)\|_\infty \leqslant \delta/2,$$

for sufficiently large n. Since $\delta > 0$ was arbitrary, the claim follows on observing that $x(\cdot), \bar{x}^n(\cdot)$ are linearly interpolated from their values at $\{t_i\}$. □

Next define $\tilde{x}^n(t), t \in I_n$, by $\tilde{x}^n(t_{m(n)}) = x(t_{m(n)})$ and

$$\dot{\tilde{x}}^n(t) = F^{y(t)}(\tilde{x}^n(t)) - \tilde{x}^n(t), t \in I_n.$$

Lemma 3.4.

$$\lim_{n \to \infty} \sup_{t \in I_n} \|\tilde{x}^n(t) - \bar{x}^n(t)\|_\infty = 0.$$

Proof. This follows easily from the Gronwall lemma. □

Let $\alpha > 0$ be as in Lemma 3.1.

Lemma 3.5. *Almost surely, the following holds. There exists an α-thick sequence $\bar{\mu}^n \in U, n \geqslant 0$, such that if $\hat{x}^n(t), t \in I_n$, is defined by $\hat{x}^n(t_{m(n)}) = x(t_{m(n)})$ and*

$$\dot{\hat{x}}^n(t) = F^{\bar{\mu}^n_t}(\hat{x}^n(t)) - \hat{x}(t), t \in I_n,$$

for $n \geqslant 1$, then

$$\lim_{n \to \infty} \sup_{t \in I_n} \|\hat{x}^n(t) - \tilde{x}^n(t)\|_\infty = 0.$$

Proof. This is immediate from Lemmas 2.3 and 3.1. □

Proof of Theorem 1.1. Let $\varepsilon > 0$. Let $b = \alpha$ above in Corollary 2.1 and pick $T = T(\varepsilon, \alpha)$ accordingly. Pick $\gamma > 0$ as in Corollary 2.2. Combining Lemmas 3.3–3.5, we have

$$\limsup_{n\to\infty}\sup_{t\in I_n} \|\hat{x}^n(t) - x(t)\|_\infty = 0 \text{ a.s.}$$

Thus $x(t_n + \cdot)$ is a (α, T, γ)-perturbation of (5) for sufficiently large n. By Corollary 2.2, it follows that $x(t) \to G^\varepsilon$. Since $\varepsilon > 0$ is arbitrary, the claim follows. □

Observe that the foregoing can be easily extended to the following relaxation of the latter half of (A2). The directed graph formed therein need not be irreducible, but each communicating class in it must correspond to elements of S which together cover I. Also, extension to nonexpansive F with respect to the weighted ∞-norm is straighforward.

4. Examples

This section sketches some important instances of fixed point problems for ∞-nonexpansive maps. A general reference for these is [1].

(i) *Shortest path problems*: Given $d+1$ locations $\{0, 1, \ldots, d\}$ and the distances $\{d_{ij}, 0 \leqslant i, j \leqslant d, i \neq j\}$ between them, the problem is to find the shortest path from location $i \neq 0$ to location 0. Letting $V(i)$ = the length of the shortest path from i to 0, one has the dynamic programming equations

$$V(i) = \min\left(d_{i0}, \min_{j\neq i,0}(d_{ij} + V(j))\right), \quad 1 \leqslant i \leqslant d.$$

Letting $V = [V(1), \ldots, V(d)]^T$, this has the form $V = F(V)$ for an ∞-nonexpansive F.

(ii) *Markov decision processes*: Consider a controlled Markov chain $\{X_n\}$ on a finite state space S, with a compact metric action space A and a continuous transition probability function $p: S \times S \times A \to [0, 1]$. The aim is to choose an A-valued sequence $\{Z_n\}$ that does not anticipate future, to minimize a suitable total expected cost. Thus

$$P(X_{n+1} = j / X_i, Z_i, i \leqslant n) = p(X_n, j, Z_n)\ \forall n.$$

Let B be a proper subset of S and $a \in (0, 1)$. Consider two cost functionals: For $k \in C(S \times A)$,

(1) cost up to a first passage time:

$$E\left[\sum_{n=0}^{\tau-1} k(X_n, Z_n)\right],$$

where $\tau = \min\{n \geqslant 0 \mid X_n \in B\}$,

(2) infinite horizon discounted cost:

$$E\left[\sum_{n=0}^{\infty} a^n k(X_n, Z_n)\right].$$

Letting $V(i)$ denote the minimum cost when $X_0 = i$, the dynamic programming equations in the two cases are, respectively,

$$V(i) = \min_u \left(k(i, u) + \sum_{j \notin B} p(i, j, u)\, V(j)\right), \quad i \in S \backslash B,$$

and

$$V(i) = \min_u \left(k(i,u) + a \sum_j p(i,j,u)\, V(j) \right), \quad i \in S.$$

Both can be cast as fixed point equations $V = F(V)$ for ∞-nonexpansive V.
(iii) *Systems of linear equations*: The problem of solving a system of linear equations $Ax = b$ can be cast as finding the fixed point of $F(x) = x - a(Ax - b), a \in (0,1)$. If $\|I - A\|_\infty \leqslant 1$, F is ∞-nonexpansive.
(iv) *Strictly convex network flow problems*: The following problem arises in network flow optimization:

$$\text{minimize} \sum_{(i,j) \in A} a_{ij}(f_{ij})$$

subject to

$$\sum_{\{j|(i,j) \in A\}} f_{ij} - \sum_{\{j|(j,i) \in A\}} f_{ji} = S_i \,\forall i \in N, b_{ij} \leqslant f_{ij} \leqslant c_{ij} \forall (i,j) \in A,$$

where $a_{ij}(\cdot)$ are strictly convex. This problem can be cast as that of finding a fixed point of a pseudononexpansive map, i.e., a map F satisfying $\|F(x) - y\|_\infty \leqslant \|x - y\|_\infty$ whenever y is a fixed point of F. (See (1), § 7.2 for details). Our analysis applies here as well.

Acknowledgements

This research was supported by a grant from the ISRO-IISc Space Technology Cell. The author would like to thank Prof. Vinod Sharma for pointing out an error in the preliminary version of this paper.

References

[1] Bertsekas D P and Tsitsiklis J N, *Parallel and distributed computation* (Prentice Hall, Englewood Cliffs) (1989)
[2] Borkar V S, Topics in Controlled Markov Chains, Pitman Research Notes in Maths, No. 240, Longman Scientific and Technical, Harlow (1991)
[3] Borkar V S and Phansalkar V V, Managing interprocessor delays in distributed recursive algorithms, *Sadhana* **19** (1994) 995–1003
[4] Borkar V S and Soumyanath K, A new analog parallel scheme for fixed point computation, Part I: Theory, preprint (1993)
[5] Hirsch M W, Convergent activation dynamics in continuous time networks, *Neural Networks* **2** (1987) 331–349
[6] Neveu J, *Discrete Parameter Martingales* (North Holland: Amsterdam) (1975)