Asynchronous SA with Differential Inclusions

Kintan Saha

November 2025

Aim

- This work combines two branches of stochastic approximation theory:
  - the framework for asynchronous SA developed by Borkar, and
  - the differential inclusion–based analysis of Benaïm et al.
- The key contribution involves an analysis that relies on assumptions which can be verified *a priori*, before running the algorithm.
- This contrasts with earlier work, where assumptions often needed to be checked *during* the execution of the algorithm, i.e., simultaneously with the evolution of the iterates.
- To demonstrate an application of their proposed method, the authors analyse a two-timescale algorithm for finding the optimal policy in an MDP under weaker assumptions than previous works.

SA with Differential Inclusions

A Marchaud map is a set–valued map $F(\cdot) : X \to 2^Y$ satisfying:

(i) $F(\cdot)$ is a **closed** (equivalently upper semicontinuous) set–valued map, i.e.

$$\mathrm{Graph}(F) = \{(x, y) : y \in F(x)\}$$

is closed.

(ii) For every $x \in X$, the set $F(x)$ is **nonempty, compact, and convex**.

(iii) There exists $c > 0$ such that for all $x \in X$,

$$\sup_{z \in F(x)} \|z\| \leq c(1 + \|x\|).$$

A differential inclusion is an evolution equation of the form:

$$\frac{dx}{dt} \in F(x),$$

where F is a Marchaud map.

SA with Differential Inclusions

A **solution** is an absolutely continuous trajectory $x(\cdot)$ such that

$$\frac{dx(t)}{dt} \in F(x(t)) \quad \text{for a.e. } t.$$

The **flow** of the differential inclusion is the set-valued map:

$$\Phi_t(x) = \{x(t) : x(\cdot) \text{ solves DI with } x(0) = x\}.$$

Asymptotic pseudo-trajectory (APT): A continuous map $y : \mathbb{R}_+ \to \mathbb{R}^K$ is an APT of Φ if for every $T > 0$,

$$\lim_{t\to\infty} \sup_{s\in[0,T]} d\big(y(t+s), \Phi_s(y(t))\big) = 0,$$

where $d(\cdot, \cdot)$ is the distance from a point to a set.

SA with Differential Inclusions

Theorem 2.3

Consider the following stochastic approximation under the assumptions:

$$x_{n+1} \in x_n + \alpha(n+1)\big[F(x_n) + V_{n+1} + d_{n+1}\big],$$

(i) (**Kushner–Clark noise condition**) For all $T > 0$,

$$\lim_{n\to\infty} \sup_{k=n+1,\ldots,m(\tau_n+T)} \left\| \sum_{i=n}^{k-1} \alpha(i+1)V_{i+1} \right\| = 0.$$

(ii) The iterates are bounded: $\sup_n \|x_n\| < \infty$.

(iii) $F(\cdot)$ is a Marchaud map.

(iv) $d_n \to 0$ and $\sup_n \|d_n\| < \infty$.

Then a linear interpolation of $\{x_n\}$ is an **asymptotic pseudo-trajectory** of the differential inclusion

$$\frac{dx}{dt} \in F(x).$$

Notation and Problem Setting

- Let $I = \{1, \dots, K\}$ index the components of $x_n \in \mathbb{R}^K$. At iteration n, let $\bar{I}_n$ be the (random) subset of components updated at time n. For each $i \in I$, the **local update counter** is
$$\nu_n(i) = \sum_{k=1}^{n} \mathbf{1}_{\{i \in \bar{I}_k\}}.$$
- Let $F_i(x) = \{y_i : y \in F(x)\}$ denote the ith coordinate of the set-valued mean field. The **asynchronous SA update** is
$$x_{n+1}(i) \in x_n(i) + \alpha(\nu_{n+1}(i))\, \mathbf{1}_{\{i \in \bar{I}_{n+1}\}} \big(F_i(x_n) + V_{n+1}(i) + d_{n+1}(i)\big).$$
- Define the **global step size** and the **relative step sizes** as:
$$\bar{\alpha}_n = \max_{i \in \bar{I}_n} \alpha(\nu_n(i)), \quad \mu_n(i) = \frac{\alpha(\nu_n(i))}{\bar{\alpha}_n} \mathbf{1}_{\{i \in \bar{I}_n\}} \in [0, 1].$$
- Let $M_n = \mathrm{diag}(\mu_n(1), \dots, \mu_n(K))$. Then the asynchronous SA with DI takes the form:
$$\boxed{x_{n+1} - x_n - \bar{\alpha}_{n+1} M_{n+1}(V_{n+1} + d_{n+1}) \;\in\; \bar{\alpha}_{n+1} M_{n+1} \cdot F(x_n)}$$

Asynchronous SA with DI

- Combining M_{n+1} with $F(x_n)$ fails because M_{n+1} is time-varying and can be zero infinitely often, making $M_{n+1}F(x_n)$ vanish even when $F(x_n) \neq 0$. Thus the limiting DI may differ from the synchronous case. To avoid this, following Borkar, we use weak limits of $\{M_n\}$, which are bounded away from zero under mild assumptions.
- For each n, define $f_n \in F(x_n)$ componentwise via

$$f_n(i)\mu_{n+1}(i) = \frac{x_{n+1}(i) - x_n(i)}{\bar{\alpha}_{n+1}} - \mu_{n+1}(i)\big(V_{n+1}(i) + d_{n+1}(i)\big),$$

  with any $f_n(i) \in F_i(x_n)$ chosen when $\mu_{n+1}(i) = 0$.
- Then the SA recursion becomes

$$x_{n+1} = x_n + \bar{\alpha}_{n+1} M_{n+1}\big(f_n + V_{n+1} + d_{n+1}\big).$$

- Introduce diagonal matrices $\widetilde{M}_n$ whose entries lie in $[\varepsilon, 1]$ for some fixed $\varepsilon > 0$; this ensures they are bounded away from zero .
- Rewrite the recursion as

$$\boxed{x_{n+1} = x_n + \bar{\alpha}_{n+1}\big[\widetilde{M}_{n+1} f_n + (M_{n+1} - \widetilde{M}_{n+1}) f_n + M_{n+1} V_{n+1} + M_{n+1} d_{n+1}\big]}$$

Asynchronous SA with DI

- Define modified noise terms $\bar{V}_{n+1} = f_n(M_{n+1} - \widetilde{M}_{n+1}) + M_{n+1}V_{n+1}$ and $\bar{d}_{n+1} = M_{n+1}d_{n+1}$, yielding
$$x_{n+1} = x_n + \bar{\alpha}_{n+1}\big(\widetilde{M}_{n+1}f_n + \bar{V}_{n+1} + \bar{d}_{n+1}\big).$$
- For $\Omega_K^\delta = \{\mathrm{diag}(\omega_1, \ldots, \omega_K) : \omega_i \in [\delta, 1]\}$, define the averaged mean field
$$\bar{F}(x) = \Omega_K^\varepsilon \cdot F(x) = \{\, Mf \,:\, M \in \Omega_K^\varepsilon,\ f \in F(x)\,\}.$$
- The recursion becomes a standard SA with set-valued mean field:
$$\boxed{x_{n+1} - x_n - \bar{\alpha}_{n+1}\bar{V}_{n+1} - \bar{\alpha}_{n+1}\bar{d}_{n+1} \ \in\ \bar{\alpha}_{n+1}\bar{F}(x_n)}.$$
- The **asynchronous timescale** is defined by
$$\bar{\tau}_0 = 0, \qquad \bar{\tau}_n = \sum_{k=1}^{n} \bar{\alpha}_k,$$
and the **interpolated trajectory** is:
$$\bar{x}(\bar{\tau}_n + s) = x_n + s\,\frac{x_{n+1} - x_n}{\bar{\alpha}_{n+1}}, \qquad s \in [0, \bar{\alpha}_{n+1}),$$

Main results

- The paper then introduces assumptions (A1)–(A5) and constructs a sequence $\{\widetilde{M}_n\}$ of diagonal matrices bounded in $[\varepsilon, 1]$ such that $\bar{\alpha}_n$ and $\bar{V}_n$ satisfy the conditions required for convergence of SA with Differential Inclusions

Theorem 3.1

Under assumptions (A1)–(A5), with probability 1, the interpolated trajectory $\bar{x}(t)$ is an asymptotic pseudo-trajectory of the differential inclusion

$$\frac{dx}{dt} \in \bar{F}(x),$$

where $\bar{F}(\cdot)$ is as defined before.

Corollary 3.2

If, for all $\varepsilon > 0$, the differential inclusion admits a global attractor $A \subset C$, and assumptions (A1)–(A5) hold, then the asynchronous iterative process converges to A almost surely.

Assumptions (A1)–(A5)

- **(A1) Boundedness.** (a) The iterates remain in a compact set $C \subset \mathbb{R}^K$. (b) The sequence $\{d_n\}$ is bounded and $d_n \to 0$.
- **(A2) Step-size conditions.** (a) $\sum_n \alpha(n) = \infty$ and $\alpha(n) \to 0$. (b) The step-size is nonincreasing and satisfies $\sup_n \alpha([xn])/\alpha(n) < A_x < \infty$ for all $x \in (0,1)$.
- **(A3) Marchaud mean field.** $F(\cdot)$ is a Marchaud map.
- **(A4) Controlled Markovian updates.** Let $\bar{I} \subset 2^I$ denote the collection of all update–sets that occur with positive probability. Let $\mathcal{F}_n$ be the sigma–algebra

$$\mathcal{F}_n := \sigma\left(\{\bar{I}_m\}_m, \{x_m\}_m, \{\nu_m(i)\}_{i,m} : m \leq n\right).$$

  Assume that for all $x \in C$ and all $\bar{I}_n, \bar{I}_{n+1} \in \bar{I}$:

  (a) $\mathbb{P}\left(\bar{I}_{n+1} = \bar{I}_{n+1} \,\middle|\, \mathcal{F}_n\right) = \mathbb{P}\left(\bar{I}_{n+1} = \bar{I}_{n+1} \,\middle|\, \bar{I}_n = \bar{I}_n,\ x_n = x\right) =: P_{(\bar{I}_n, \bar{I}_{n+1})}(x)$.

  (b) For all $x \in C$, the transition probabilities $P_{(\bar{I}_n, \bar{I}_{n+1})}(x)$ form an aperiodic, irreducible Markov chain on $\bar{I}$, and for every coordinate $i \in I$ there exists some update–set $\mathcal{I} \in \bar{I}$ with $i \in \mathcal{I}$.

  (c) The map $x \mapsto P_{(\bar{I}_n, \bar{I}_{n+1})}(x)$ is Lipschitz continuous.

Assumptions

- **(A5) Noise conditions.** Assume that for all n, V_{n+1} and $\bar{I}_{n+1}$ are uncorrelated given $\mathcal{F}_n$, and either:
  (a) There exists $q \geq 2$ such that

$$\sum_n \alpha(n)^{1+q/2} < \infty, \qquad \sup_n \mathbb{E}\left[\|V_n\|^q\right] < \infty.$$

  (b) The coordinates $V_n(i)$ are independent for $i \neq j$, and for the inner product $\langle a, b\rangle = \sum_k a_k b_k$, there exists a constant $\Gamma > 0$ such that for all $\theta \in \mathbb{R}^K$:

$$\mathbb{E}[\exp\{\langle \theta, V_{n+1}\rangle\} \mid \mathcal{F}_n] \leq \exp\left\{\frac{\Gamma}{2}\|\theta\|^2\right\}.$$

  In addition,

$$\sum_n e^{-c/\alpha(n)} < \infty \qquad \text{for each } c > 0.$$

SA with Differential Inclusions

We consider the following stochastic approximation under the assumptions:

$$x_{n+1} \in x_n + \bar{\alpha}(n+1)\left[\bar{F}(x_n) + \bar{V}_{n+1} + \bar{d}_{n+1}\right],$$

(i) (**Kushner–Clark noise condition**) For all $T > 0$,

$$\lim_{n\to\infty} \sup_{k=n+1,\ldots,\bar{m}(\bar{\tau}_n+T)} \left\| \sum_{i=n}^{k-1} \bar{\alpha}(i+1)\bar{V}_{i+1} \right\| = 0.$$

(ii) The iterates are bounded: $\sup_n \|x_n\| < \infty$.

(iii) $\bar{F}(\cdot)$ is a Marchaud map.

(iv) $\bar{d}_n \to 0$ and $\sup_n \|\bar{d}_n\| < \infty$.

Then a linear interpolation of $\{x_n\}$ is an **asymptotic pseudo-trajectory** of the differential inclusion

$$\frac{dx}{dt} \in \bar{F}(x).$$

Methodology

- Now to verify the Kushner–Clark noise condition for the Differential Inclusion SA, we must show that for every $T > 0$,

$$\lim_{n\to\infty} \sup_{k=n+1,\ldots,\bar{m}(\bar{\tau}_n+T)} \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1}\, \bar{V}_{i+1} \right\| = 0,$$

where $\bar{V}_{i+1} = f_i(M_{i+1} - \widetilde{M}_{i+1}) + M_{i+1}V_{i+1}$ and $\bar{m}(t) := sup\{k \geq 0; t \geq \bar{\tau}_k\}$.

- Using this definition and the triangle inequality, for fixed $T > 0$ we obtain

$$\sup_k \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1}\, \bar{V}_{i+1} \right\| \leq \sup_k \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} M_{i+1} V_{i+1} \right\| + \sup_k \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} f_i\big(M_{i+1} - \widetilde{M}_{i+1}\big) \right\|,$$

for $k = n+1, \ldots, \bar{m}(\bar{\tau}_n + T)$. We will show that *each* of these two terms converges to 0 almost surely, which establishes the Kushner–Clark condition for $\{\bar{V}_n\}$.

Proof Sketch for Theorem 3.1

Lemma 3.3

Assume (A2)(b), (A4), and (A5). Then with probability 1, for all $T > 0$,

$$\lim_{n\to\infty} \sup_{k=n+1,\ldots,\bar{m}(\bar{\tau}_n+T)} \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} M_{i+1} V_{i+1} \right\| = 0.$$

Lemma 3.6

Almost surely under (A1)(a), (A2), and (A4), for every $T > 0$,

$$\lim_{n\to\infty} \sup_{k=n+1,\ldots,\bar{m}(\bar{\tau}_n+T)} \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} f_i \left(M_{i+1} - \widetilde{M}_{i+1} \right) \right\| = 0.$$

- Lemmas 3.3 and 3.6 together verify condition (i).
- Assumption (A1)(a) immediately gives condition (ii).
- Under (A1)–(A5), the averaged map $\bar{F}(\cdot)$ is Marchaud, proving (iii).
- Finally, assumption (A1)(b) is equivalent to condition (iv).

Two-Timescale Asynchronous SA

- The paper extends the single-timescale asynchronous SA framework to two coupled processes updated on different timescales:

$$x_n \in \mathbb{R}^K, \qquad y_n \in \mathbb{R}^L.$$

- Let I index the coordinates of x_n and J index those of y_n. Define the sets of feasible asynchronous update–subsets and nth iteration updated coordinates as::

$$\bar{I} \subset 2^I, \quad \bar{J} \subset 2^J, \quad \bar{I}_n \in \bar{I}, \quad \bar{J}_n \in \bar{J}..$$

- Define the local counters:

$$\nu_n(i) = \sum_{k=1}^{n} \mathbf{1}_{\{i \in \bar{I}_k\}}, \qquad \phi_n(j) = \sum_{k=1}^{n} \mathbf{1}_{\{j \in \bar{J}_k\}}.$$

- Noise sequences: $V_n, d_n \in \mathbb{R}^K, \quad U_n, e_n \in \mathbb{R}^L,$ where V_n, U_n are martingale differences and $d_n, e_n \to 0$.
- Stepsizes: $\alpha(n), \ \gamma(n) > 0$ decreasing, with $\alpha(n)$ governing the x–updates and $\gamma(n)$ the y–updates.

Two-Timescale Asynchronous SA

- Set-valued mean fields $F : \mathbb{R}^K \times \mathbb{R}^L \to 2^{\mathbb{R}^K}$, $G : \mathbb{R}^K \times \mathbb{R}^L \to 2^{\mathbb{R}^L}$, with coordinate maps $F_i(x, y) = \{z_i : z \in F(x, y)\}$, $G_j(x, y) = \{z_j : z \in G(x, y)\}$.
- Asynchronous two-timescale updates:

$$x_{n+1}(i) - x_n(i) \in \alpha(\nu_{n+1}(i))\,\mathbf{1}_{\{i \in \bar{I}_{n+1}\}}\big(F_i(x_n, y_n) + V_{n+1}(i) + d_{n+1}(i)\big),$$
$$y_{n+1}(j) - y_n(j) \in \gamma(\phi_{n+1}(j))\,\mathbf{1}_{\{j \in \bar{J}_{n+1}\}}\big(G_j(x_n, y_n) + U_{n+1}(j) + e_{n+1}(j)\big).$$

- Global/relative step sizes:

$$\bar{\alpha}_n = \max_{i \in \bar{I}_n} \alpha(\nu_n(i)), \quad \mu_n(i) = \frac{\alpha(\nu_n(i))}{\bar{\alpha}_n}\mathbf{1}_{\{i \in \bar{I}_n\}},$$
$$\bar{\gamma}_n = \max_{j \in \bar{J}_n} \gamma(\phi_n(j)), \quad \sigma_n(j) = \frac{\gamma(\phi_n(j))}{\bar{\gamma}_n}\mathbf{1}_{\{j \in \bar{J}_n\}}.$$

Matrices: $M_n = \mathrm{diag}(\mu_n(1), \dots, \mu_n(K))$, $N_n = \mathrm{diag}(\sigma_n(1), \dots, \sigma_n(L))$.

$$\boxed{\begin{aligned} x_{n+1} - x_n - \bar{\alpha}_{n+1} M_{n+1}(V_{n+1} + d_{n+1}) \ &\in \ \bar{\alpha}_{n+1} M_{n+1} F(x_n, y_n), \\ y_{n+1} - y_n - \bar{\gamma}_{n+1} N_{n+1}(U_{n+1} + e_{n+1}) \ &\in \ \bar{\gamma}_{n+1} N_{n+1} G(x_n, y_n) \end{aligned}}$$

Assumptions

The paper extends the single–timescale assumptions to the two–timescale case.

- **(B1)**
  - (a) There exist compact sets $C \subset \mathbb{R}^K$, $D \subset \mathbb{R}^L$ such that $x_n \in C$, $y_n \in D$ for all n.
  - (b) The perturbation sequences $\{d_n\}$ and $\{e_n\}$ are bounded and satisfy $d_n, e_n \to 0$.
- **(B2)** For both $a(n) = \alpha(n)$ and $a(n) = \gamma(n)$:
  - (a) $\sum_n a(n) = \infty$ and $a(n) \to 0$.
  - (b) For all $x \in (0,1)$, $\sup_n a([xn])/a(n) < A_x < \infty$ and $a(n)$ is nonincreasing.
  - (c) Two-timescale separation: $\alpha(n)/\gamma(n) \to 0$.
- **(B3)**
  - (a) $F(\cdot,\cdot) : C \times D \to C$ is upper semicontinuous, and for each $y \in D$, $F(\cdot, y)$ is a Marchaud map.
  - (b) $G(\cdot,\cdot) : C \times D \to D$ is a Marchaud map.
- **(B6)** $\bar{G}(x,y) := \Omega_L^{\varepsilon} \cdot G(x,y)$. For all $x \in C$ and every $\varepsilon > 0$, the differential inclusion

$$\frac{dy}{dt} \in \bar{G}(x,y)$$

  has a unique global attractor $\Lambda(x)$. Moreover, $\Lambda : \mathbb{R}^K \to \mathbb{R}^L$ is bounded, continuous, and single–valued for all $x \in C$.

Assumptions

- **(B4)** Define $\bar{H} \subset \bar{I} \times \bar{J}$ as the set of update pairs with positive simultaneous probability. Let $z_n = (x_n, y_n)$ and $\mathcal{F}_n = \sigma(\{\bar{H}_m\}_m, \{z_m\}_m, \{\nu_m(i)\}_{i,m}, \{\phi_m(j)\}_{j,m} : m \leq n)$.
  (a) For all $z \in C \times D$,
  $$\mathbb{P}(\bar{H}_{n+1} \mid \mathcal{F}_n) = \mathbb{P}(\bar{H}_{n+1} \mid \bar{H}_n, z_n = z) =: Q(\bar{H}_n, \bar{H}_{n+1})(z)$$
  (b) For all $z \in C \times D$, $Q(\cdot,\cdot)(z)$ is aperiodic and irreducible on $\bar{H}$, and for each $i \in I$, $j \in J$, there exist $\bar{H}, \bar{H}' \in \bar{H}$ with $i \in \bar{H}$ and $j \in \bar{H}'$.
  (c) The map $z \mapsto Q(\bar{H}_n, \bar{H}_{n+1})(z)$ is Lipschitz continuous.
- **(B5)** Separately for both $(a(n), W_n, \bar{Z}_n) = (\alpha(n), V_n, \bar{I}_n)$ and $(\gamma(n), U_n, \bar{J}_n)$, assume W_{n+1} and $\bar{Z}_{n+1}$ are uncorrelated given $\mathcal{F}_n$, and either:
  (a) For some $q \geq 2$,
  $$\sum_n a(n)^{1+q/2} < \infty, \qquad \sup_n \mathbb{E}\left[\|W_n\|^q\right] < \infty.$$
  (b) $W_n(i)$ is independent of $W_n(j)$ for $i \neq j$. With $\langle a, b\rangle = \sum_k a_k b_k$, assume there exists $\Gamma > 0$ such that for all $\theta \in \mathbb{R}^K/\mathbb{R}^L$,
  $$\mathbb{E}[\exp\{\langle \theta, W_{n+1}\rangle\} \mid \mathcal{F}_n] \leq \exp\{\tfrac{\Gamma}{2}\|\theta\|^2\},$$
  and $\sum_n e^{-c/a(n)} < \infty, \qquad \forall c > 0.$

Convergence of fast timescale

Similar to the single–timescale setting, the coupled two–timescale recursion can be rewritten in a compact form. The updates are:

$$x_{n+1} = x_n + \bar{\alpha}_{n+1} M_{n+1}\big(f_n + V_{n+1} + d_{n+1}\big), \qquad y_{n+1} = y_n + \bar{\gamma}_{n+1} N_{n+1}\big(g_n + U_{n+1} + e_{n+1}\big)$$

for $f_n \in F(x_n, y_n)$ and $g_n \in G(x_n, y_n)$.
Let 0^K denote the zero vector in $\mathbb{R}^K$, and define

$$z_n := \begin{pmatrix} x_n \\ y_n \end{pmatrix}, \qquad \Gamma_n := \begin{pmatrix} M_n \\ N_n \end{pmatrix}, \qquad \zeta_n := \begin{pmatrix} \frac{\bar{\alpha}_n}{\bar{\gamma}_n} V_n \\ U_n \end{pmatrix},$$

$$\kappa_{n+1} := \begin{pmatrix} \frac{\bar{\alpha}_n}{\bar{\gamma}_n}(f_n + d_{n+1}) \\ e_{n+1} \end{pmatrix}, \qquad \Psi(z_n) := \begin{pmatrix} 0^K \\ G(z_n) \end{pmatrix}.$$

Then the coupled process can be expressed in the single iterative form

$$\boxed{z_{n+1} - z_n - \bar{\gamma}_{n+1}\Gamma_{n+1}\big(\zeta_{n+1} + \kappa_{n+1}\big) \;\in\; \bar{\gamma}_{n+1}\Gamma_{n+1} \cdot \Psi(z_n).}$$

Convergence of fast timescale

Lemma 4.2

Under (B2)(b), (B4), (B5), with prob. 1,

$$\lim_{n\to\infty} \sup_{k=n+1,\dots,\bar{m}_\gamma(\bar{\rho}_n+T)} \left\| \sum_{j=n}^{k-1} \bar{\gamma}_{j+1}\Gamma_{j+1}\zeta_{j+1} \right\| = 0.$$

Assumption (B4) allows us to construct $\widetilde{\Gamma}_n \in \Omega^{\varepsilon}_{K+L}$ with the same weak limit as Γ_n, and let $\bar{z}(t)$ be the linear interpolation of the iterates. Define the modified mean field

$$\bar{\Psi}(z) := \Omega^{\varepsilon}_{K+L} \cdot \Psi(z).$$

Lemma 4.3

Under (B1)–(B6), with prob. 1, the interpolated trajectory $\bar{z}(t)$ is an asymptotic pseudo–trajectory of

$$\frac{dz}{dt} \in \bar{\Psi}(z).$$

Convergence of fast timescale

Let $\bar{x}(t)$, $\bar{y}(t)$ be the interpolations of the iterates.

Corollary 4.4

Under (B1)–(B6), with prob. 1,

$$(\bar{x}(t), \bar{y}(t)) \to \{(x, \Lambda(x)) : x \in C\} \qquad (t \to \infty).$$

Since the fast variable tracks $y = \Lambda(x)$, define the reduced mean field

$$F^{\Lambda}(x) := F(x, \Lambda(x)) \subset \mathbb{R}^{K}.$$

Under (B3)(a) and (B6), F^{Λ} is a Marchaud map. Following the single–timescale construction, define the averaged field

$$\bar{F}^{\Lambda}(x) := \Omega_K^{\varepsilon} \cdot F^{\Lambda}(x).$$

Convergence of slow timescale

Jointly Perturbed Solution (DI 4.7)

A continuous function $z : \mathbb{R}^+ \to C$ is a jointly perturbed solution of

$$\frac{dz}{dt} \in \bar{F}^{\Lambda}(x),$$

if

(i) z is absolutely continuous;

(ii) $\frac{dz}{dt} - V(t) - d(t) \in \bar{F}^{\Lambda}_{\delta(t)}(z(t))$ for a.e. $t > 0$, where $d(t)$, $\delta(t) \to 0$ and

$$\bar{F}^{\Lambda}_{\delta}(x) = \left\{ f \in C : \|x' - x\| \leq \delta,\ \|f' - f\| \leq \delta,\ f' \in \bar{F}^{\Lambda}(x') \right\};$$

(iii) $V(t)$ is locally integrable and

$$\lim_{t\to\infty} \sup_{0\leq\nu\leq T} \left\| \int_t^{t+\nu} V(s)\, ds \right\| = 0.$$

Convergence of Slow Timescale

Lemma 4.6

Under (B1)–(B6), any jointly perturbed solution of the differential inclusion

$$\frac{dx}{dt} \in \bar{F}^{\Lambda}(x)$$

is an asymptotic pseudo–trajectory of the flow induced by this DI.

Theorem 4.7

Under (B1)–(B6), with probability 1, the interpolated slow process $\bar{x}(t)$ is a jointly perturbed solution and hence an asymptotic pseudo–trajectory of

$$\frac{dx}{dt} \in \bar{F}^{\Lambda}(x).$$

Corollary 4.8

If for all $\varepsilon > 0$ the DI above admits a global attractor $A \subset C$, then the slow component $\{x_n\}$ of the two–timescale process converges almost surely to A.

Two–Timescale Learning in an MDP

Setup.

- We operate in a discounted MDP with unknown transition kernel and rewards.
- At each step the agent observes (s_n, a_n, r_n, s_{n+1}).
- The goal is to learn an optimal stationary policy.

Two–timescale update structure.

- *Fast timescale*:
$$Q_{n+1}(s,a) = Q_n(s,a) + \gamma(\phi_{n+1}(s,a))\, \mathbf{1}_{\{(s,a)=(s_{n+1},a_{n+1})\}} \Big[R_{n+1} + \beta V_n(s_{n+2}) - Q_n(s,a) \Big]$$
- *Slow timescale*: $\pi_{n+1}(s) = \pi_n(s) + \mu(\nu_{n+1}(s))\, \mathbf{1}_{\{s=s_{n+1}\}} \Big[b_s(Q_{n+1}) - \pi_n(s) \Big]$
- Stepsizes satisfy $\alpha(n)/\gamma(n) \to 0$ (critic faster than actor).

Asynchronous setting.

- States and actions are visited irregularly.
- Updates use local counters $\nu_n(s)$, $\phi_n(s,a)$.
- The asynchronous Markov visitation process satisfies (B4)–(B6).

Main Convergence Result

Theorem (Convergence of Two–Timescale MDP Learner). Under assumptions (B1)–(B6) and persistent exploration, the two–timescale asynchronous algorithm satisfies:

$$\boxed{\pi_n \longrightarrow \pi^* \qquad \text{almost surely}}$$

where π^* is an optimal stationary policy of the MDP.

Proof sketch.

- **Fast timescale:** Value estimates track the differential inclusion associated with the Bellman operator with $y = \Lambda(x)$ fixed. Lemmas 4.2–4.4 show the value estimator is an asymptotic pseudo–trajectory of its DI and converges to the unique fixed point Q^{π}.
- **Slow timescale:** With the critic converging, the policy update follows the DI $\dot{x} \in \bar{F}^{\Lambda}(x)$, whose attractor corresponds to the optimal responses to the limiting value function.
- **Chain of inclusions:** Lemma 4.6 and Theorem 4.7 imply the slow interpolated trajectory is an asymptotic pseudo–trajectory of its DI. Corollary 4.8 ensures convergence to the global attractor set, which is the set of optimal policies.
- **Conclusion:** Combining fast and slow DI limits yields $\pi_n \to \pi^*$ **almost surely**.

Remarks

- The MDP learning algorithm cannot be analyzed using Borkar's asynchronous SA theory, since Borkar requires globally Lipschitz mean fields F, G, while the policy update here uses a discontinuous *best–response* operator.
- None of the proofs in the paper actually require $x_n \in C$ for a compact C; the weaker assumption $\sup_n \|x_n\| < \infty$ suffices throughout.
- The authors' key advantage is that their assumptions are verifiable *a priori*. In particular, (A4) fully specifies the controlled Markov chain governing update sets, and can be checked before execution. Borkar instead requires the visitation condition

$$\liminf_{n\to\infty} \frac{\nu_n(i)}{n} > 0 \quad \text{a.s.},$$

which cannot be known beforehand. Moreover, (A4) itself implies this property (Lemma A.1), avoiding the need to assume it.

Thank You!

Any Questions?