Distributed Optimization in Adaptive Networks: Appendix

Ciamac Moallemi
Electrical Engineering
Stanford University
Stanford, CA 94305
ciamac@stanford.edu

Benjamin Van Roy
Management Science and Engineering
and Electrical Engineering
Stanford University
Stanford, CA 94305
bvr@stanford.edu

November 18, 2003

1 Markov Decision Processes

Consider a Markov chain $(w(k), a(k))$ defined for $k = 0, 1, \ldots$ and with $w(k) \in \mathbb{W}$, $a(k)$ in $\mathbb{A}$, where $\mathbb{W}$ and $\mathbb{A}$ are finite sets representing the system state space and the action space, respectively. The transition probabilities are defined by the function

$$P_\theta(w', a', w, a) = \Pr\left\{ w(k+1) = w, a(k+1) = a \,|\, w(k) = w', a(k) = a' \right\}.$$

Here, $\theta \in \mathbb{R}^N$ is a vector of policy parameters.

We will make the following assumption regarding the dynamics.

Assumption 1.1. *For all θ, the Markov chain $(w(k))$ is ergodic (aperiodic, irreducible).*

While the system is in state $w \in \mathbb{W}$ and action $a \in \mathbb{A}$ is applied, a reward $r(w, a)$ is accrued. We will use the shorthand $r(k) = r(w(k), a(k))$. Given Assumption 1.1, we can define the long term average reward by

$$\begin{aligned} \lambda(\theta) &= \lim_{K\to\infty} \frac{1}{K} \operatorname{E}\left[\sum_{k=0}^{K-1} r(k)\right] \\ &= \sum_{w\in\mathbb{W}, a\in\mathbb{A}} \eta_\theta(w, a) r(w, a), \end{aligned}$$

where $n_\theta(w,a)$ is the steady-state distribution corresponding to the transition function $P_\theta(w',a',w,a)$.

Define the differential reward function

$$q_\theta(w,a) = \lim_{K\to\infty} \mathrm{E}\left[\sum_{k=0}^{K-1} \left(r(w(k),a(k)) - \lambda(\theta)\right) \middle| w(0)=w, a(0)=a\right].$$

The following result provides a crucial expression for the gradient of $\lambda(\theta)$. It is important in that it does not rely on terms of the form $\nabla_\theta \eta_\theta(w,a)$, which would be difficult to estimate over finite sample paths. It is a standard result in Markov decision process theory, see [3], for example, for a proof.

Theorem 1.1. *Assume that $P_\theta(w',a',w,a)$ is continuously differentiable with respect to θ. Then,*

$$\nabla_\theta \lambda(\theta) = \sum_{w\in\mathbb{W}, a\in\mathbb{A}} \sum_{w'\in\mathbb{W}, a'\in\mathbb{A}} \eta_\theta(w',a') \nabla_\theta P_\theta(w',a',w,a) q_\theta(w,a). \tag{1.1}$$

2 Network Structure

Assume the network has n components. Corresponding to each component i, there is a subset $\mathbb{W}_i \in \mathbb{W}$. At the kth epoch, there are a set of control actions $a_1(k) \in \mathbb{A}_1, \ldots, a_n(k) \in \mathbb{A}_n$, where each $\mathbb{A}_1, \ldots, \mathbb{A}_n$ is a finite set. We will denote the entire action vector $(a_1(k), \ldots, a_n(k))$ as $a(k) \in \mathbb{A} = \mathbb{A}_1 \times \cdots \times \mathbb{A}_n$. Actions are governed by a set of policies $\pi^1_{\theta_1}, \ldots, \pi^n_{\theta_n}$, where the policy $\pi^i_{\theta_i}$ at component i is parameterized by a vector $\theta_i \in \mathbb{R}^{N_i}$. Each ith action process transitions only if the state $w(k)$ is an element of $\mathbb{W}_i$. At the time of transition, the probability that $a_i(k)$ becomes any $a_i \in \mathbb{A}_i$ is given by $\pi^i_{\theta_i}(a_i|w(k))$. Hence, the corresponding action sequence evolves according to

$$a_i(k) = \begin{cases} a_i' & \text{with probability } \pi^i_{\theta_i}(a_i'|w(k)), \text{ if } w(k) \in \mathbb{W}_i, \\ a_i(k-1) & \text{otherwise.} \end{cases}$$

The state transitions depend on the prior state and action vector. In particular, there is a transition kernel P that defines the state dynamics:

$$\Pr\left\{w(k) = w | w(k-1) = w', a(k-1) = a'\right\} = P(w',a',w).$$

Hence, if $\theta = (\theta_1, \ldots, \theta_n)$, we have

$$P_\theta(w',a',w,a) = P(w',a',w) \prod_{i:w\in\mathbb{W}_i} \pi^i_{\theta_i}(a_i|w) \prod_{i:w\notin\mathbb{W}_i} \mathbf{1}_{\{a_i'=a_i\}}. \tag{2.1}$$

Finally, we will assume that the reward is an average of rewards occurring at each component, that is

$$r(w,a) = \frac{1}{n}\sum_{i=1}^{n} r_i(w,a).$$

We will use the shorthand $r_i(k) = r_i(w(k), a(k))$.

We will make the following assumption regarding the policies.

Assumption 2.1. *For all i and every $w \in \mathbb{W}_i$, $a_i \in \mathbb{A}_i$, $\pi^i_{\theta_i}(a_i|w)$ is a continuously differentiable function of θ_i. Further, for every i, there exists a bounded function $L_i(w, a_i, \theta)$ such that for all $w \in \mathbb{W}_i, a_i \in \mathbb{A}_i$,*

$$\nabla_{\theta_i}\pi^i_{\theta_i}(a_i|w) = \pi^i_{\theta_i}(a_i|w)L_i(w,a_i,\theta).$$

The latter part of the assumption is satisfied, for example, if there exists a constant $\epsilon > 0$ such that for each i,$w \in \mathbb{W}_i$,$a_i \in \mathbb{A}_i$,

$$\text{either } \forall\theta_i, \pi^i_{\theta_i}(a_i|w) = 0 \text{ or } \forall\theta_i, \pi^i_{\theta_i}(a_i|w) \geq \epsilon.$$

Without loss of generality, we will assume that $\pi^i_{\theta_i}(a_i|w) > 0$, and hence define a bound L by

$$\sup_{i,\theta_i,w\in\mathbb{W}_i,a_i\in\mathbb{A}_i}\left\|\frac{\nabla_{\theta_i}\pi^i_{\theta_i}(a_i|w)}{\pi^i_{\theta_i}(a_i|w)}\right\| < L.$$

In this framework, the gradient expression of Theorem 1.1 becomes significantly simpler.

Theorem 2.1. *For all i,*

$$\nabla_{\theta_i}\lambda(\theta) = \sum_{w\in\mathbb{W}_i,a\in\mathbb{A}} \eta_\theta(w,a)\frac{\nabla_{\theta_i}\pi^i_{\theta_i}(a_i|w)}{\pi^i_{\theta_i}(a_i|w)}q_\theta(w,a).$$

Proof. Examining (2.1), it is clear that

$$\nabla_{\theta_i}P_\theta(w',a',w,a) = P_\theta(w',a',w,a)\frac{\nabla_{\theta_i}\pi^i_{\theta_i}(a_i|w)}{\pi^i_{\theta_i}(a_i|w)}\mathbf{1}_{\{w\in\mathbb{W}_i\}}.$$

Combining with Theorem 1.1, we have

$$\begin{aligned}
\nabla_{\theta_i}\lambda(\theta) &= \sum_{\substack{w,a\\w',a'}} \eta_\theta(w',a')P_\theta(w',a',w,a)\frac{\nabla_{\theta_i}\pi^i_{\theta_i}(a_i|w)}{\pi^i_{\theta_i}(a_i|w)}\mathbf{1}_{\{w\in\mathbb{W}_i\}}q_\theta(w,a)\\
&= \sum_{w,a}\frac{\nabla_{\theta_i}\pi^i_{\theta_i}(a_i|w)}{\pi^i_{\theta_i}(a_i|w)}\mathbf{1}_{\{w\in\mathbb{W}_i\}}q_\theta(w,a)\sum_{w',a'}\eta_\theta(w',a')P_\theta(w',a',w,a)\\
&= \sum_{w\in\mathbb{W}_i,a\in\mathbb{A}}\eta_\theta(w,a)\frac{\nabla_{\theta_i}\pi^i_{\theta_i}(a_i|w)}{\pi^i_{\theta_i}(a_i|w)}q_\theta(w,a).
\end{aligned}$$

□

3 Centralized Gradient Estimation

For $\beta \in (0,1]$, define the eligibility vector

$$z_i^\beta(k) = \sum_{\ell=0}^{k} \beta^{k-\ell} \frac{\nabla_{\theta_i} \pi_{\theta_i}^i(a_i(\ell)|w(\ell))}{\pi_{\theta_i}^i(a_i(\ell)|w(\ell))} \mathbf{1}_{\{w(\ell)\in\mathbb{W}_i\}} \tag{3.1}$$

$$= \beta z_i^\beta(k-1) + \frac{\nabla_{\theta_i} \pi_{\theta_i(k)}^i(a_i(k)|w(k))}{\pi_{\theta_i(k)}^i(a_i(k)|w(k))} \mathbf{1}_{\{w(k)\in\mathbb{W}_i\}}. \tag{3.2}$$

We can define a centralized estimate of the gradient $\nabla_{\theta_i}\lambda(\theta)$ by

$$\bar{\chi}_i(k) = r(k) z_i^\beta(k),$$

where we are using the shorthand $r(k) = r(w(k), a(k))$.

Define $\nabla_i(k)$ as shorthand for

$$\frac{\nabla_{\theta_i} \pi_{\theta_i}^i(a_i(k)|w(k))}{\pi_{\theta_i}^i(a_i(k)|w(k))} \mathbf{1}_{\{w(k)\in\mathbb{W}_i\}}.$$

The following lemma will be useful in subsequent analysis.

Lemma 3.1. *If* $\ell < k$, $\mathrm{E}[\nabla_i(k)|\mathcal{F}_\ell] = 0$.

Proof. Note that for $\ell < k$,

$$\begin{aligned}
\mathrm{E}[\nabla_i(k)|\mathcal{F}_\ell] &= \sum_{w\in\mathbb{W}_i} \sum_{a_i\in\mathbb{A}_i} \Pr\{w(k)=w|\,\mathcal{F}_\ell\}\, \pi_{\theta_i}^i(a_i|w) \left[\frac{\nabla_{\theta_i}\pi_{\theta_i}^i(a_i|w)}{\pi_{\theta_i}^i(a_i|w)}\right] \\
&= \sum_{w\in\mathbb{W}_i} \Pr\{w(k)=w|\,\mathcal{F}_\ell\} \sum_{a_i\in\mathbb{A}_i} \nabla_{\theta_i}\pi_{\theta_i}^i(a_i|w) \\
&= \sum_{w\in\mathbb{W}_i} \Pr\{w(k)=w|\,\mathcal{F}_\ell\}\, \nabla_{\theta_i} \left(\sum_{a_i\in\mathbb{A}_i} \pi_{\theta_i}^i(a_i|w)\right) \\
&= \sum_{w\in\mathbb{W}_i} \Pr\{w(k)=w|\,\mathcal{F}_\ell\}\, \nabla_{\theta_i}(1) \\
&= 0.
\end{aligned}$$

□

We will now establish convergence of long term averages of the discounted gradient estimator. Note that a stronger result is proved in [1], however the following is sufficient for our purposes.

Theorem 3.1. *For any* i *and* $0 < \beta < 1$*,*

$$\lim_{K\to\infty} \frac{1}{K} \mathrm{E}\left[\sum_{k=0}^{K-1} \bar{\chi}_i(k)\right] = \sum_{w\in\mathbb{W}_i, a\in\mathbb{A}} \eta_\theta(w,a) \frac{\nabla_{\theta_i}\pi^i_{\theta_i}(a_i|w)}{\pi^i_{\theta_i}(a_i|w)} q^\beta_\theta(w,a),$$

where $q^\beta_\theta(w,a)$ *is the discounted differential reward function*

$$q^\beta_\theta(w,a) = \lim_{K\to\infty} \mathrm{E}\left[\sum_{k=0}^{K-1} \beta^k \left(r(w(k),a(k)) - \lambda(\theta)\right) \middle| w(0)=w, a(0)=a\right].$$

Further,

$$\lim_{\beta\uparrow 1} \lim_{K\to\infty} \frac{1}{K} \mathrm{E}\left[\sum_{k=0}^{K-1} \bar{\chi}_i(k)\right] = \nabla_{\theta_i}\lambda(\theta).$$

Proof. Note that

$$\begin{aligned}
\frac{1}{K} \mathrm{E}\left[\sum_{k=0}^{K-1} \bar{\chi}_i(k)\right] &= \frac{1}{K} \mathrm{E}\left[\sum_{\ell=0}^{K-1} \nabla_i(\ell) \sum_{k=\ell}^{K-1} \beta^{k-\ell} r(k)\right] \\
&= \frac{1}{K} \mathrm{E}\left[\sum_{\ell=0}^{K-1} \nabla_i(\ell) \sum_{k=\ell}^{K-1} \beta^{k-\ell} \left(r(k) - \lambda(\theta)\right)\right] \\
&= \frac{1}{K} \mathrm{E}\left[\sum_{\ell=0}^{K-1} \nabla_i(\ell) q^\beta_\theta(w(\ell), a(\ell), K-\ell)\right],
\end{aligned}$$

where we use the fact the $\mathrm{E}[\nabla_i(\ell)] = 0$, from Lemma 3.1, and where

$$q^\beta_\theta(w,a,K) = \mathrm{E}\left[\sum_{k=0}^{K-1} \beta^k \left(r(w(k),a(k)) - \lambda(\theta)\right) \middle| w(0)=w, a(0)=a\right].$$

It is clear the $q^\beta_\theta(w,a,K) \to q^\beta_\theta(w,a)$ as $K\to\infty$, then, since $\nabla_i(\ell)$ is bounded, it follows that

$$\begin{aligned}
\lim_{K\to\infty} \frac{1}{K} \mathrm{E}\left[\sum_{k=0}^{K-1} \bar{\chi}_i(k)\right] &= \lim_{K\to\infty} \frac{1}{K} \mathrm{E}\left[\sum_{\ell=0}^{K-1} \nabla_i(\ell) q^\beta_\theta(w(\ell), a(\ell))\right] \\
&= \sum_{w\in\mathbb{W},_i a\in\mathbb{A}} \eta_\theta(w,a) \frac{\nabla_{\theta_i}\pi^i_{\theta_i}(a_i|w)}{\pi^i_{\theta_i}(a_i|w)} q^\beta_\theta(w,a),
\end{aligned}$$

where the last step follows since $(w(\ell), a(\ell))$ is ergodic (Assumption 1.1). The balance of the result follows from the fact that $\lim_{\beta \uparrow 1} q_\theta^\beta(w, a) = q_\theta(w, a)$. $\square$

4 Distributed Gradient Estimation

Consider the following gradient estimator:

$$\chi_i(k) = z_i^\beta(k) \frac{1}{n} \sum_{j=1}^{n} \sum_{\ell=0}^{k} d_{ij}^\alpha(\ell, k) r_j(\ell), \tag{4.1}$$

Here, the random variables $\{d_{ij}^\alpha(\ell, k)\}$, with parameter $\alpha \in (0, 1)$, represent an arrival process describing the communication of rewards across the network. Indeed, $d_{ij}^\alpha(\ell, k)$ is the fraction of the reward $r_j(\ell)$ at component j that is learned by component i at time $k \geq \ell$. We will assume the arrival process satisfies the following conditions.

Assumption 4.1. *For each i, j, ℓ, and $\alpha \in (0, 1)$, the process $\{d_{ij}^\alpha(\ell, k) | k = \ell, \ell + 1, \ell + 2, \ldots\}$ satisfies:*

1. *$d_{ij}^\alpha(\ell, k)$ is $\mathcal{F}_k$-measurable.*
2. *There exists a scalar $\gamma \in (0, 1)$ and a random variable c_ℓ such that for all $k \geq \ell$,*

$$\left| \frac{d_{ij}^\alpha(\ell, k)}{(1 - \alpha)\alpha^{k-\ell}} - 1 \right| < c_\ell \gamma^{k-\ell},$$

   with probability 1. Further, we require that the distribution of c_ℓ given $\mathcal{F}_\ell$ depend only on $(w(\ell), a(\ell))$, and that there exist a constant $\bar{c}$ such that

$$\mathrm{E}\left[c_\ell | \, w(\ell) = w, a(\ell) = a \right] < \bar{c} < \infty,$$

   with probability 1 for all initial conditions $w \in \mathbb{W}$ and $a \in \mathbb{A}$.

3. *The distribution of $\{d_{ij}^\alpha(\ell, k) | k = \ell, \ell + 1, \ldots\}$ given $\mathcal{F}_\ell$ depends only on $w(\ell)$ and $a(\ell)$.*

Note that from Assumption 4.1(2), it is clear that $\sum_{k=\ell}^\infty d_{ij}^\alpha(\ell, k)$ converges absolutely with probability 1. Further, we have

$$\begin{aligned} \left| \sum_{k=\ell}^{\infty} \left(d_{ij}^\alpha(\ell, k) - (1 - \alpha)\alpha^{k-\ell} \right) \right| &< \sum_{k=\ell}^{\infty} c_\ell (1 - \alpha) \alpha^{k-\ell} \gamma^{k-\ell} \\ &= \frac{c_\ell (1 - \alpha)}{1 - \alpha\gamma}. \end{aligned}$$

Hence, with probability 1,

$$\lim_{\alpha\uparrow 1}\sum_{k=\ell}^{\infty} d_{ij}^{\alpha}(\ell,k) = \lim_{\alpha\uparrow 1}\sum_{k=\ell}^{\infty}(1-\alpha)\alpha^{k-\ell} = 1. \tag{4.2}$$

5 Relation to Centralized Gradient Estimation

For convenience, define $R = \max_{i,a,w} |r_i(w,a)|$. The following lemma will be useful throughout this analysis.

Lemma 5.1. *There exists constants C and $\eta \in (0,1)$ such that, for all k, l, and any functions g and f,*

$$\begin{aligned}|\mathrm{E}\left[g(w(\ell),a(\ell))f(w(k),a(k))\right] - \mathrm{E}\left[g(w(\ell),a(\ell))\right]\mathrm{E}\left[f(w(k),a(k))\right]| \\ \le \max_{w,a}|f(w,a)|\max_{w,a}|g(w,a)|C\eta^{|k-\ell|}.\end{aligned}$$

In particular, for an arbitrary function f,

$$\|\mathrm{E}\left[f(w(\ell),a(\ell)\nabla_i(k)\right]\| \le \max_{w,a}|f(w,a)|LC\eta^{|k-\ell|},$$

Proof. The first statement follows immediately from Assumption 1.1. The second statement follows from the first once we observe (from Lemma 3.1) that

$$\mathrm{E}[\nabla_i(k)] = 0.$$

□

Lemma 5.2. *For each i, j, $k \ge \ell$, $\alpha \in (0,1)$ and $\beta \in (0,1)$,*

$$\mathrm{E}\left[\left\|z_i^{\beta}(k)d_{ij}^{\alpha}(\ell,k)\right\|\bigg|\,\mathcal{F}_\ell\right] < \frac{(1-\alpha)(1+\bar{c})L\alpha^{k-\ell}}{1-\beta}.$$

Proof. From Assumption 4.1(2),

$$|d_{ij}^{\alpha}(\ell,k)| < (1-\alpha)(1+c_\ell)\alpha^{k-\ell}.$$

Then,

$$\begin{aligned}\left\|z_i^{\beta}(k)d_{ij}^{\alpha}(\ell,k)\right\| &\le (1-\alpha)(1+c_\ell)L\alpha^{k-\ell}\sum_{u=0}^{k}\beta^{k-u} \\ &< \frac{(1-\alpha)(1+c_\ell)L\alpha^{k-\ell}}{1-\beta}.\end{aligned}$$

The result follows after taking a conditional expectation. □

Let

$$\hat{z}_{ij}^{\alpha\beta}(\ell, K) = \mathrm{E}\left[\left.\sum_{k=\ell}^{K-1} z_i^{\beta}(k) d_{ij}^{\alpha}(\ell, k)\right| \mathcal{F}_\ell\right].$$

By Lemma 5.2, for $\alpha \in (0,1)$ and $\beta \in (0,1)$, $\{\hat{z}_{ij}^{\alpha\beta}(\ell, K) | K = \ell, \ell+1, \ell+2, \ldots\}$ is a Cauchy sequence, and therefore,

$$\hat{z}_{ij}^{\alpha\beta}(\ell) = \lim_{K\to\infty} \hat{z}_{ij}^{\alpha\beta}(\ell, K),$$

is well-defined and finite. The following lemma follows immediately.

Lemma 5.3. *For any i and j, $\alpha \in (0,1)$, and $\beta \in (0,1)$,*

$$\lim_{K\to\infty} \left\| \frac{1}{K} \mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell) \left(\hat{z}_{ij}^{\alpha\beta}(\ell, K) - \hat{z}_{ij}^{\alpha\beta}(\ell)\right)\right] \right\| = 0.$$

Lemma 5.4. *For any i, ℓ, and $\alpha \in (0,1)$, $\lim_{K\to\infty} \hat{z}_{ij}^{\alpha 1}(\ell, K)$ is well-defined. Further, if we define $\hat{z}_{ij}^{\alpha 1}(\ell) = \lim_{K\to\infty} \hat{z}_{ij}^{\alpha 1}(\ell, K)$, then for any j,*

$$\limsup_{\alpha\uparrow 1} \limsup_{K\to\infty} \left\| \frac{1}{K} \mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell) \left(\hat{z}_{ij}^{\alpha 1}(\ell) - z_i^1(\ell)\right)\right] \right\| = 0.$$

Proof. Note that

$$\begin{aligned}
&\hat{z}_{ij}^{\alpha 1}(\ell, K) \\
&= \mathrm{E}\left[\left.\sum_{k=\ell}^{K-1} \sum_{s=0}^{k} \nabla_i(s) d_{ij}^{\alpha}(\ell, k)\right| \mathcal{F}_\ell\right] \\
&= \mathrm{E}\left[\left.\sum_{s=0}^{\ell} \nabla_i(s) \sum_{k=\ell}^{K-1} d_{ij}^{\alpha}(\ell, k)\right| \mathcal{F}_\ell\right] + \mathrm{E}\left[\left.\sum_{s=\ell+1}^{K-1} \nabla_i(s) \sum_{k=s}^{K-1} d_{ij}^{\alpha}(\ell, k)\right| \mathcal{F}_\ell\right] \\
&= G_{ij}^{\alpha}(\ell, K) + H_{ij}^{\alpha}(\ell, K).
\end{aligned}$$

For the term $G_{ij}^{\alpha}(\ell, K)$, note that

$$\lim_{K\to\infty} G_{ij}^{\alpha}(\ell, K) = z_i^1(\ell) \lim_{K\to\infty} f_{ij}^{\alpha}(w(\ell), a(\ell), K - \ell),$$

where, using Assumption 4.1(3), we define

$$f_{ij}^{\alpha}(w, a, K) = \mathrm{E}\left[\left.\sum_{k=0}^{K-1} d_{ij}^{\alpha}(0, k)\right| w(0) = w, a(0) = a\right].$$

Note that for $J < K$, from Assumption 4.1(2),

$$\begin{aligned}\left|f_{ij}^{\alpha}(w,a,K) - f_{ij}^{\alpha}(w,a,J)\right| &\leq (1-\alpha)(1+\bar{c})\sum_{k=J}^{K-1}\alpha^k \\ &\leq (1+\bar{c})\alpha^J.\end{aligned}$$

Hence, for $\alpha \in (0,1)$, $\{f_{ij}^{\alpha}(w,a,K)|K = 1,2,\ldots\}$ is a Cauchy sequence, and we can define the limit

$$f_{ij}^{\alpha}(w,a) = \lim_{K\to\infty} f_{ij}^{\alpha}(w,a,K).$$

Further, the following limit exists,

$$\lim_{K\to\infty} \mathrm{E}\left[G_{ij}^{\alpha}(\ell,K)\middle|\,\mathcal{F}_\ell\right] = z_i^1(\ell) f_{ij}^{\alpha}(w(\ell),a(\ell)).$$

For the term $H_{ij}^{\alpha}(\ell,K)$, note that for $J < K$,

$$\begin{aligned}&\left\|\mathrm{E}\left[H_{ij}^{\alpha}(\ell,K) - H_{ij}^{\alpha}(\ell,J)\middle|\,\mathcal{F}_\ell\right]\right\| \\ &= \left\|\mathrm{E}\left[\sum_{s=J}^{K-1}\nabla_i(s)\sum_{k=s}^{K-1}d_{ij}^{\alpha}(\ell,k) + \sum_{s=\ell+1}^{J-1}\nabla_i(s)\sum_{k=J}^{K-1}d_{ij}^{\alpha}(\ell,k)\middle|\,\mathcal{F}_\ell\right]\right\| \\ &\leq L(1-\alpha)(1+\bar{c})\left(\sum_{s=J}^{K-1}\sum_{k=s}^{K-1}\alpha^{k-\ell} + \sum_{s=\ell+1}^{J-1}\sum_{k=J}^{K-1}\alpha^{k-\ell}\right) \\ &\leq L(1+\bar{c})\left(\sum_{s=J}^{K-1}\alpha^{s-\ell} + \sum_{s=\ell+1}^{J-1}\alpha^{J-\ell}\right) \\ &\leq L(1+\bar{c})\left(\frac{\alpha^J}{1-\alpha} + (J-\ell+1)\alpha^{J-\ell}\right).\end{aligned}$$

Hence, $\{H_{ij}^{\alpha}(\ell,K)|K = \ell+1,\ell+2,\ldots\}$ is a Cauchy sequence. Then, we can define

$$\hat{z}_{ij}^{\alpha 1}(\ell) = \lim_{K\to\infty}\hat{z}_{ij}^{\alpha 1}(\ell,K).$$

To establish the balance of the result, note that

$$\begin{aligned}
&\left\| \frac{1}{K} \mathsf{E} \left[\sum_{\ell=0}^{K-1} r_j(\ell) \left(\hat{z}_{ij}^{\alpha 1}(\ell) - z_i^1(\ell) \right) \right] \right\| \\
&= \left\| \frac{1}{K} \mathsf{E} \left[\sum_{\ell=0}^{K-1} r_j(\ell) \lim_{M\to\infty} \left(G_{ij}^{\alpha}(\ell, M) + H_{ij}^{\alpha}(\ell, M) - z_i^1(\ell) \right) \right] \right\| \\
&\leq \left\| \frac{1}{K} \mathsf{E} \left[\sum_{\ell=0}^{K-1} r_j(\ell) \left(1 - f_{ij}^{\alpha}(w(\ell), a(\ell)) \right) z_i^1(\ell) \right] \right\| \\
&\quad + \left\| \frac{1}{K} \mathsf{E} \left[\sum_{k=0}^{K-1} r_j(\ell) \lim_{M\to\infty} H_{ij}^{\alpha}(\ell, M) \right] \right\| \\
&= \textbf{(A)} + \textbf{(B)}.
\end{aligned}$$

For term **(A)**, note that

$$\begin{aligned}
&\left\| \frac{1}{K} \mathsf{E} \left[\sum_{\ell=0}^{K-1} r_j(\ell) \left(1 - f_{ij}^{\alpha}(w(\ell), a(\ell)) \right) z_i^1(\ell) \right] \right\| \\
&= \left\| \frac{1}{K} \mathsf{E} \left[\sum_{\ell=0}^{K-1} \sum_{u=0}^{\ell} r_j(\ell) \left(1 - f_{ij}^{\alpha}(w(\ell), a(\ell)) \right) \nabla_i(u) \right] \right\| \\
&\leq \frac{RLC}{K} \sum_{\ell=0}^{K-1} \sum_{u=0}^{\ell} \eta^{\ell-u} \max_{w\in\mathbb{W}, a\in\mathbb{A}} \left| 1 - f_{ij}^{\alpha}(w, a) \right| \\
&\leq \frac{RLC}{1-\eta} \max_{w\in\mathbb{W}, a\in\mathbb{A}} \left| 1 - f_{ij}^{\alpha}(w, a) \right|.
\end{aligned}$$

Note that this bound is independent of K, and, by the Dominated Convergence Theorem and (4.2), $\lim_{\alpha\uparrow 1} f_{ij}^{\alpha}(w, a) = 1$, hence the **(A)** term vanishes.

For term **(B)**, note that for $s > \ell$, $\mathrm{E}\left[\nabla_i(s)\middle|\,\mathcal{F}_\ell\right] = 0$ from Lemma 3.1. Hence,

$$\begin{aligned}
&\left\|\frac{1}{K}\,\mathrm{E}\left[\sum_{k=0}^{K-1} r_j(\ell)\lim_{M\to\infty} H_{ij}^{\alpha}(\ell,K)\right]\right\| \\
&= \left\|\frac{1}{K}\,\mathrm{E}\left[\sum_{k=0}^{K-1} r_j(\ell)\lim_{M\to\infty}\mathrm{E}\left[\sum_{s=\ell+1}^{M-1}\nabla_i(s)\sum_{k=s}^{K-1} d_{ij}^{\alpha}(\ell,k)\middle|\,\mathcal{F}_\ell\right]\right]\right\| \\
&= \left\|\mathrm{E}\left[\sum_{s=\ell+1}^{K-1}\nabla_i(s)\sum_{k=s}^{K-1}\left(d_{ij}^{\alpha}(\ell,k)-(1-\alpha)\alpha^{k-\ell}\right)\middle|\,\mathcal{F}_\ell\right]\right\| \\
&\le \sum_{s=\ell+1}^{K-1} L\sum_{k=s}^{K-1}\mathrm{E}\left[c_\ell(1-\alpha)\alpha^{k-\ell}\gamma^{k-\ell}\middle|\,\mathcal{F}_\ell\right] \\
&\le \bar{c}(1-\alpha)L\sum_{s=\ell+1}^{K-1}\frac{\alpha^{s-\ell}\gamma^{s-\ell}}{1-\alpha\gamma} \\
&\le \bar{c}(1-\alpha)L\frac{\alpha\gamma}{(1-\alpha\gamma)^2}.
\end{aligned}$$

Note that this bound is independent of K and tends to 0 as $\alpha \uparrow 1$. Hence, term **(B)** vanishes and the result is established. □

Because the limit is well-defined, we extend our definition of $\hat{z}_{ij}^{\alpha\beta}(\ell)$ to the case of $\beta = 1$:

$$\hat{z}_{ij}^{\alpha 1}(\ell) = \lim_{K\to\infty}\hat{z}_{ij}^{\alpha 1}(\ell,K).$$

Lemma 5.5. *For any i and j,*

$$\limsup_{\beta\uparrow 1}\limsup_{K\to\infty}\left\|\frac{1}{K}\,\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell)\left(z_i^1(\ell)-z_i^{\beta}(\ell)\right)\right]\right\| = 0.$$

Proof. We have

$$\begin{aligned}
\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell)\left(z_i^1(\ell)-z_i^{\beta}(\ell)\right)\right] &= \mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell)\sum_{k=0}^{\ell}(1-\beta^{\ell-k})\nabla_i(k)\right] \\
&= \sum_{\ell=0}^{K-1}\sum_{k=0}^{\ell}(1-\beta^{\ell-k})\,\mathrm{E}\left[r_j(\ell)\nabla_i(k)\right]
\end{aligned}$$

From Lemma 5.1,

$$\left\|\mathrm{E}\left[r_j(\ell)\nabla_i(k)\right]\right\| \le RLC\eta^{\ell-k}.$$

It follows that

$$\begin{aligned}\left\|\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell)\left(z_i^1(\ell)-z_i^\beta(\ell)\right)\right]\right\| &\leq RLC\sum_{\ell=0}^{K-1}\sum_{k=0}^{\ell}\left(1-\beta^{\ell-k}\right)\eta^{\ell-k}\\ &\leq KRLC\left(\frac{1}{1-\eta}-\frac{1}{1-\beta\eta}\right),\end{aligned}$$

The result follows. □

Lemma 5.6. *For any i, j, and $\alpha \in (0,1)$,*

$$\limsup_{\beta\uparrow 1}\limsup_{K\to\infty}\left\|\frac{1}{K}\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell)\left(\hat{z}_{ij}^{\alpha\beta}(\ell)-\hat{z}_{ij}^{\alpha 1}(\ell)\right)\right]\right\| = 0.$$

Proof. Note that

$$\begin{aligned}&\frac{1}{K}\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell)\left(\hat{z}_{ij}^{\alpha\beta}(\ell)-\hat{z}_{ij}^{\alpha 1}(\ell)\right)\right]\\ &= \frac{1}{K}\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell)\lim_{M\to\infty}\mathrm{E}\left[\sum_{k=\ell}^{M-1}\left(z_i^\beta(k)-z_i^1(k)\right)d_{ij}^\alpha(\ell,k)\middle|\mathcal{F}_\ell\right]\right]\\ &= \frac{1}{K}\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell)\lim_{M\to\infty}\mathrm{E}\left[\sum_{k=\ell}^{M-1}\left(\beta^{k-\ell}z_i^\beta(\ell)-z_i^1(\ell)\right)d_{ij}^\alpha(\ell,k)\middle|\mathcal{F}_\ell\right]\right]\\ &\quad +\frac{1}{K}\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell)\right.\\ &\qquad \left.\lim_{M\to\infty}\mathrm{E}\left[\sum_{k=\ell}^{M-1}\left(z_i^\beta(k)-\beta^{k-\ell}z_i^\beta(\ell)-z_i^1(k)+z_i^1(\ell)\right)d_{ij}^\alpha(\ell,k)\middle|\mathcal{F}_\ell\right]\right]\\ &= \textbf{(A)}+\textbf{(B)}.\end{aligned}$$

Consider term **(A)**. From Assumption 4.1(3), we can define

$$g_{ij}^{\alpha\beta}(w,a,M) = \mathrm{E}\left[\sum_{k=0}^{M-1}\beta^k d_{ij}^\alpha(0,k)\middle| w(0)=w, a(0)=a\right].$$

By Assumption 4.1(2), for $\alpha \in (0,1)$ and $\beta \in [0,1]$, and for $J < K$

$$\begin{aligned}
&\left| g_{ij}^{\alpha\beta}(w,a,K) - g_{ij}^{\alpha\beta}(w,a,J) \right| \\
&\quad \leq \quad \mathrm{E}\left[(1-\alpha)(1+c_0) \sum_{k=J}^{K-1} \alpha^k \beta^k \middle| w(0)=w, a(0)=a \right] \\
&\quad \leq \quad \frac{(1-\alpha)(1+\bar{c})\alpha^J \beta^J}{1-\alpha\beta}.
\end{aligned}$$

Hence, $\{g_{ij}^{\alpha\beta}(w,a,M) | M = 1,2,\ldots\}$ is a Cauchy sequence, and we can define the limit

$$g_{ij}^{\alpha\beta}(w,a) = \lim_{M\to\infty} g_{ij}^{\alpha\beta}(w,a,M).$$

Then, term **(A)** becomes

$$\begin{aligned}
&\frac{1}{K}\left\| \mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell) \lim_{M\to\infty} \mathrm{E}\left[\sum_{k=\ell}^{M-1} \left(\beta^{k-\ell} z_i^{\beta}(\ell) - z_i^1(\ell) \right) d_{ij}^{\alpha}(\ell,k) \middle| \mathcal{F}_\ell \right] \right] \right\| \\
&= \quad \frac{1}{K}\left\| \mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell) \left(g_{ij}^{\alpha\beta}(w(\ell),a(\ell)) z_i^{\beta}(\ell) - g_{ij}^{\alpha 1}(w(\ell),a(\ell)) z_i^1(\ell) \right) \right] \right\| \\
&= \quad \frac{1}{K}\left\| \mathrm{E}\left[\sum_{\ell=0}^{K-1} \sum_{u=0}^{\ell} r_j(\ell) \left(\beta^{\ell-u} g_{ij}^{\alpha\beta}(w(\ell),a(\ell)) - g_{ij}^{\alpha 1}(w(\ell),a(\ell)) \right) \nabla_i(u) \right] \right\|
\end{aligned}$$

Note that

$$\begin{aligned}
&\left| \beta^{\ell-u} g_{ij}^{\alpha\beta}(w(\ell),a(\ell)) - g_{ij}^{\alpha 1}(w(\ell),a(\ell)) \right| \\
&\quad \leq \quad \lim_{M\to\infty} \mathrm{E}\left[(1-\alpha)(1+c_0) \sum_{k=0}^{M-1} (1-\beta^{k+\ell-u})\alpha^k \middle| w(0)=w, a(0)=a \right] \\
&\quad \leq \quad (1-\alpha)(1+\bar{c}) \left(\frac{1}{1-\alpha} - \frac{\beta^{\ell-u}}{1-\alpha\beta} \right)
\end{aligned}$$

From Lemma 5.1,

$$\begin{aligned}
&\left\| \mathrm{E}\left[r_j(\ell) \left(\beta^{\ell-u} g_{ij}^{\alpha\beta}(w(\ell),a(\ell)) - g_{ij}^{\alpha 1}(w(\ell),a(\ell)) \right) \nabla_i(u) \right] \right\| \\
&\quad \leq \quad RLC(1-\alpha)(1+\bar{c})\eta^{\ell-u} \left(\frac{1}{1-\alpha} - \frac{\beta^{\ell-u}}{1-\alpha\beta} \right).
\end{aligned}$$

Applying this to term **(A)**,

$$
\begin{aligned}
&\frac{1}{K}\left\|\mathrm{E}\left[\sum_{\ell=0}^{K-1}\sum_{u=0}^{\ell} r_j(\ell)\left(\beta^{\ell-u} g_{ij}^{\alpha\beta}(w(\ell),a(\ell)) - g_{ij}^{\alpha 1}(w(\ell),a(\ell))\right)\nabla_i(u)\right]\right\| \\
&\quad\leq \frac{RLC(1-\alpha)(1+\bar{c})}{K}\sum_{\ell=0}^{K-1}\sum_{u=0}^{\ell}\eta^{\ell-u}\left(\frac{1}{1-\alpha}-\frac{\beta^{\ell-u}}{1-\alpha\beta}\right) \\
&\quad\leq \frac{RLC(1-\alpha)(1+\bar{c})}{K}\sum_{\ell=0}^{K-1}\left(\frac{1}{(1-\alpha)(1-\eta)}-\frac{1}{(1-\alpha\beta)(1-\eta\beta)}\right) \\
&\quad= RLC(1-\alpha)(1+\bar{c})\left(\frac{1}{(1-\alpha)(1-\eta)}-\frac{1}{(1-\alpha\beta)(1-\eta\beta)}\right),
\end{aligned}
$$

which is a constant over K and vanishes as $\beta \uparrow 1$.

We are left with term **(B)**. Note that

$$
\begin{aligned}
&\mathrm{E}\left[r_j(\ell)\lim_{M\to\infty}\mathrm{E}\left[\sum_{k=\ell}^{M-1}\left(z_i^{\beta}(k)-\beta^{k-\ell}z_i^{\beta}(\ell)-z_i^{1}(k)+z_i^{1}(\ell)\right)d_{ij}^{\alpha}(\ell,k)\,\middle|\,\mathcal{F}_\ell\right]\right] \\
&\quad\leq RL\,\mathrm{E}\left[\lim_{M\to\infty}\mathrm{E}\left[\sum_{k=\ell}^{M-1}(1-\alpha)(1+c_\ell)\alpha^{k-\ell}\sum_{u=\ell+1}^{k}(1-\beta^{k-u})\,\middle|\,\mathcal{F}_\ell\right]\right] \\
&\quad\leq RL\,\mathrm{E}\left[\lim_{M\to\infty}\sum_{k=\ell}^{M-1}(1-\alpha)(1+\bar{c})\alpha^{k-\ell}\left(k-\ell-\frac{1-\beta^{k-\ell}}{1-\beta}\right)\right] \\
&\quad\leq RL\,\mathrm{E}\left[(1-\alpha)(1+\bar{c})\left(\frac{\alpha}{(1-\alpha)^2}-\frac{1}{1-\beta}\left(\frac{1}{1-\alpha}-\frac{1}{1-\alpha\beta}\right)\right)\right] \\
&\quad= RL(1-\alpha)(1+\bar{c})\left(\frac{\alpha}{(1-\alpha)^2}-\frac{\alpha}{(1-\alpha)(1-\alpha\beta)}\right),
\end{aligned}
$$

which, is a constant independent of ℓ and goes to 0 as $\beta \uparrow 1$. The result follows. □

Lemma 5.7. *For all i, and $\alpha \in (0,1)$, $\beta \in (0,1)$,*

$$
\lim_{K\to\infty}\frac{1}{K}\,\mathrm{E}\left[\sum_{k=0}^{K-1}\chi_i(k)\right],
$$

exists.

Proof. We have

$$
\begin{aligned}
&\frac{1}{K}\,\mathrm{E}\left[\sum_{k=0}^{K-1}\chi_i(k)\right]\\
&= \frac{1}{K}\,\mathrm{E}\left[\sum_{j=1}^{n}\sum_{\ell=0}^{K-1}r_j(\ell)\sum_{k=\ell}^{K-1}z_i^{\beta}(k)d_{ij}^{\alpha}(\ell,k)\right]\\
&= \frac{1}{K}\,\mathrm{E}\left[\sum_{j=1}^{n}\sum_{\ell=0}^{K-1}r_j(\ell)z_i^{\beta}(\ell)\sum_{k=\ell}^{K-1}\beta^{k-\ell}d_{ij}^{\alpha}(\ell,k)\right]\\
&\quad +\frac{1}{K}\,\mathrm{E}\left[\sum_{j=1}^{n}\sum_{\ell=0}^{K-1}r_j(\ell)\sum_{k=\ell}^{K-1}\left(z_i^{\beta}(k)-\beta^{k-\ell}z_i^{\beta}(\ell)\right)d_{ij}^{\alpha}(\ell,k)\right]\\
&= \textbf{(A)}+\textbf{(B)}.
\end{aligned}
$$

We will first examine term **(A)**. Define

$$
f_{ij}^{\alpha\beta}(w,a,K)=\mathrm{E}\left[\sum_{k=0}^{K-1}\beta^k d_{ij}^{\alpha}(0,k)\,\middle|\,w(0)=w,a(0)=a\right].
$$

By Assumption 4.1(2), for $J<K$,

$$
\begin{aligned}
&\left|f_{ij}^{\alpha\beta}(w,a,K)-f_{ij}^{\alpha\beta}(w,a,J)\right|\\
&= \left|\mathrm{E}\left[\sum_{k=J}^{K-1}\beta^k d_{ij}^{\alpha}(0,k)\,\middle|\,w(0)=w,a(0)=a\right]\right|\\
&\le \left|\mathrm{E}\left[(1-\alpha)(1+c_0)\sum_{k=J}^{K-1}\beta^k\alpha^k\,\middle|\,w(0)=w,a(0)=a\right]\right|\\
&\le \frac{(1-\alpha)(1+\bar{c})\alpha^J\beta^J}{(1-\alpha\beta)}.
\end{aligned}
$$

Hence, $\{f_{ij}^{\alpha\beta}(w,a,K)|K=1,2,\ldots\}$ is a Cauchy sequence, and we can define the limit

$$
f_{ij}^{\alpha\beta}(w,a)=\lim_{K\to\infty}f_{ij}^{\alpha\beta}(w,a,K).
$$

Hence, we can define a constant

$$
C_{ij}^{\alpha\beta}=\sup_{w\in\mathbb{W},a\in\mathbb{A},K>0}\left|f_{ij}^{\alpha\beta}(w,a,K)\right|.
$$

Define

$$g_{ij}^{\alpha\beta}(w,a,K) = \mathrm{E}\left[\sum_{\ell=0}^{K-1} \beta^\ell r_j(\ell) f_{ij}^{\alpha\beta}(w(\ell),a(\ell),K-\ell) \middle| w(0)=w, a(0)=a\right].$$

Then, for $J<K$,

$$\begin{aligned}\left|g_{ij}^{\alpha\beta}(w,a,K) - g_{ij}^{\alpha\beta}(w,a,J)\right| &\leq 2C_{ij}^{\alpha\beta} R \sum_{\ell=J}^{K} \beta^\ell \\ &\leq \frac{2C_{ij}^{\alpha\beta} R \beta^J}{1-\beta}.\end{aligned}$$

Hence, $\{g_{ij}^{\alpha\beta}(w,a,K) | K=1,2,\ldots\}$ is a Cauchy sequence, and we can define the limit

$$g_{ij}^{\alpha\beta}(w,a) = \lim_{K\to\infty} g_{ij}^{\alpha\beta}(w,a,K).$$

Since $\mathbb{W}$ and $\mathbb{A}$ are finite, this convergence is uniform over w and a.

Returning to term **(A)**, note that using Assumption 4.1(3),

$$\begin{aligned}&\frac{1}{K}\mathrm{E}\left[\sum_{j=1}^{n}\sum_{\ell=0}^{K-1} r_j(\ell) z_i^\beta(\ell) \sum_{k=\ell}^{K-1} \beta^{k-\ell} d_{ij}^\alpha(\ell,k)\right] \\ &= \frac{1}{K}\sum_{j=1}^{n}\mathrm{E}\left[\sum_{\ell=0}^{K-}\sum_{u=0}^{\ell} r_j(\ell)\beta^{\ell-u}\nabla_i(u) \sum_{k=\ell}^{K-1} \beta^{k-\ell} d_{ij}^\alpha(\ell,k)\right] \\ &= \frac{1}{K}\sum_{j=1}^{n}\mathrm{E}\left[\sum_{u=0}^{K-1}\nabla_i(u)\sum_{\ell=u}^{K-1} r_j(\ell)\beta^{\ell-u} f_{ij}^{\alpha\beta}(w(\ell),a(\ell),K-\ell)\right] \\ &= \frac{1}{K}\sum_{j=1}^{n}\mathrm{E}\left[\sum_{u=0}^{K-1}\nabla_i(u) g_{ij}^{\alpha\beta}(w(u),a(u),K-\ell)\right]\end{aligned}$$

Since $\nabla_i(u)$ is bounded, we have

$$\lim_{K\to\infty}\frac{1}{K}\left\|\mathrm{E}\left[\sum_{u=0}^{K-1}\nabla_i(u)\left(g_{ij}^{\alpha\beta}(w(u),a(u)) - g_{ij}^{\alpha\beta}(w(u),a(u),K-u)\right)\right]\right\| = 0.$$

Yet, since $(w(\ell),a(\ell))$ is ergodic, the limit

$$\lim_{K\to\infty}\frac{1}{K}\sum_{j=1}^{n}\mathrm{E}\left[\sum_{u=0}^{K-1}\nabla_i(u) g_{ij}^{\alpha\beta}(w(u),a(u))\right],$$

exists, hence the limit of term **(A)** exists as $K \to \infty$.

We are left with term **(B)**. Note that

$$\begin{aligned}
&\frac{1}{K}\,\mathrm{E}\left[\sum_{j=1}^{n}\sum_{\ell=0}^{K-1} r_j(\ell) \sum_{k=\ell}^{K-1}\left(z_i^{\beta}(k) - \beta^{k-\ell} z_i^{\beta}(\ell)\right) d_{ij}^{\alpha}(\ell,k)\right] \\
&\quad = \frac{1}{K}\sum_{j=1}^{n}\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell) \sum_{k=\ell}^{K-1} d_{ij}^{\alpha}(\ell,k) \sum_{u=\ell+1}^{k} \beta^{k-u}\nabla_i(u)\right] \\
&\quad = \frac{1}{K}\sum_{j=1}^{n}\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell) h_{ij}^{\alpha\beta}(w(\ell), a(\ell), K-\ell)\right],
\end{aligned}$$

where

$$h_{ij}^{\alpha\beta}(w,a,K) = \mathrm{E}\left[\sum_{k=0}^{K-1} d_{ij}^{\alpha}(0,k) \sum_{u=1}^{k} \beta^{k-u}\nabla_i(u) \,\middle|\, w(0)=w, a(0)=a\right].$$

Then, for $J < K$,

$$\begin{aligned}
&\left\|h_{ij}^{\alpha\beta}(w,a,K) - h_{ij}^{\alpha\beta}(w,a,J)\right\| \\
&\quad \leq \mathrm{E}\left[L(1-\alpha)(1+c_0)\sum_{k=J}^{K-1}\alpha^k \sum_{u=1}^{k}\beta^{k-u} \,\middle|\, w(0)=w, a(0)=a\right] \\
&\quad \leq \frac{L(1-\alpha)(1+\bar{c})\alpha^J}{(1-\alpha)(1-\beta)}.
\end{aligned}$$

Hence, $\{h_{ij}^{\alpha\beta}(w,a,K) | K = 1, 2, \ldots\}$ is a Cauchy sequence, and we can define the limit

$$h_{ij}^{\alpha\beta}(w,a) = \lim_{K\to\infty} h_{ij}^{\alpha\beta}(w,a,K).$$

Then, we have

$$\lim_{K\to\infty}\frac{1}{K}\left\|\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell)\left(h_{ij}^{\alpha\beta}(w(\ell),a(\ell)) - h_{ij}^{\alpha\beta}(w(\ell),a(\ell),K-\ell)\right)\right]\right\| = 0.$$

Yet, since $(w(\ell), a(\ell))$ is ergodic, the limit

$$\lim_{K\to\infty}\frac{1}{K}\sum_{j=1}^{n}\mathrm{E}\left[\sum_{\ell=0}^{K-1} r_j(\ell) h_{ij}^{\alpha\beta}(w(\ell), a(\ell))\right],$$

exists, hence the limit of term **(B)** exists as $K \to \infty$. $\square$

Theorem 5.1. *Holding θ fixed, for all i, $\alpha \in (0,1)$, and $\beta \in (0,1)$, define*

$$\nabla^{\alpha\beta}_{\theta_i}\lambda(\theta) = \lim_{K\to\infty} \frac{1}{K} \operatorname{E}\left[\sum_{k=0}^{K-1} \chi_i(k)\right]$$

exists. Further,

$$\limsup_{\alpha\uparrow 1} \limsup_{\beta\uparrow 1} \left\| \nabla^{\alpha\beta}_{\theta_i}\lambda(\theta) - \nabla_{\theta_i}\lambda(\theta) \right\| = 0.$$

Proof. From Theorem 3.1, it suffices to prove that

$$\limsup_{\alpha\uparrow 1} \limsup_{\beta\uparrow 1} \lim_{K\to\infty} \mathcal{L}^{\alpha\beta}_i(K) = 0,$$

where

$$\mathcal{L}^{\alpha\beta}_i(K) = \left\| \frac{1}{K} \operatorname{E}\left[\sum_{k=0}^{K-1} \chi_i(k)\right] - \frac{1}{K} \operatorname{E}\left[\sum_{k=0}^{K-1} \overline{\chi}_i(k)\right] \right\|.$$

Note that from Lemma 5.7 and Theorem 3.1, $\lim_{K\to\infty} \mathcal{L}^{\alpha\beta}_i(K)$ exists when $\alpha \in (0,1)$ and $\beta \in (0,1)$.

We have

$$\begin{aligned}
&\limsup_{\beta\uparrow 1} \lim_{K\to\infty} \mathcal{L}^{\alpha\beta}_i(K) \\
&= \limsup_{\beta\uparrow 1} \lim_{K\to\infty} \left\| \frac{1}{nK} \operatorname{E}\left[\sum_{j=1}^{n}\sum_{k=0}^{K-1} r_j(\ell)\left(\hat{z}^{\alpha\beta}_{ij}(\ell,K) - z^{\beta}_i(\ell)\right)\right]\right\| \\
&\le \limsup_{\beta\uparrow 1} \limsup_{K\to\infty} \left\| \frac{1}{nK} \operatorname{E}\left[\sum_{j=1}^{n}\sum_{k=0}^{K-1} r_j(\ell)\left(\hat{z}^{\alpha\beta}_{ij}(\ell,K) - \hat{z}^{\alpha\beta}_{ij}(\ell)\right)\right]\right\| \\
&\quad + \limsup_{\beta\uparrow 1} \limsup_{K\to\infty} \left\| \frac{1}{nK} \operatorname{E}\left[\sum_{j=1}^{n}\sum_{k=0}^{K-1} r_j(\ell)\left(\hat{z}^{\alpha\beta}_{ij}(\ell) - \hat{z}^{\alpha 1}_{ij}(\ell)\right)\right]\right\| \\
&\quad + \limsup_{\beta\uparrow 1} \limsup_{K\to\infty} \left\| \frac{1}{nK} \operatorname{E}\left[\sum_{j=1}^{n}\sum_{k=0}^{K-1} r_j(\ell)\left(\hat{z}^{\alpha 1}_{ij}(\ell) - z^{1}_i(\ell)\right)\right]\right\| \\
&\quad + \limsup_{\beta\uparrow 1} \limsup_{K\to\infty} \left\| \frac{1}{nK} \operatorname{E}\left[\sum_{j=1}^{n}\sum_{k=0}^{K-1} r_j(\ell)\left(z^{1}_i(\ell) - z^{\beta}_i(\ell)\right)\right]\right\| \\
&= \textbf{(A)} + \textbf{(B)} + \textbf{(C)} + \textbf{(D)}.
\end{aligned}$$

From Lemma 5.3, term **(A)** equals 0. From Lemma 5.6, term **(B)** equals 0. From Lemma 5.5, term **(D)** equals 0. Hence, taking a limit as $\alpha \uparrow 1$,

$$\begin{aligned} 0 \leq \lim_{\alpha \uparrow 1} \limsup_{\beta \uparrow 1} \lim_{K \to \infty} \mathcal{L}_i^{\alpha\beta}(K) & \\ \leq \quad & \limsup_{\alpha \uparrow 1} \limsup_{\beta \uparrow 1} \lim_{K \to \infty} \mathcal{L}_i^{\alpha\beta}(K) \\ \leq \quad & \limsup_{\alpha \uparrow 1} \limsup_{K \to \infty} \left\| \frac{1}{nK} \mathrm{E}\left[\sum_{j=1}^{n} \sum_{k=0}^{K-1} r_j(\ell) \left(\hat{z}_{ij}^{\alpha 1}(\ell) - z_i^1(\ell) \right) \right] \right\| \\ = \quad & 0, \end{aligned}$$

where we use Lemma 5.4.

□

6 Communication Protocol

In this section, we will describe a simple protocol that allows communication of rewards in a fashion that satisfies the requirements of Assumption 4.1. This protocol communicates the rewards across the network over time using a distributed averaging procedure.

In order to motivate our protocol, consider a different problem. Imagine each component i in the network is given a real value R_i. Our goal is to design an asynchronous distributed protocol through which each node will obtain the average

$$\overline{R} = \frac{1}{n} \sum_{i=1}^{n} R_i.$$

To do this, define the vector $Y(0) \in \mathbb{R}^n$ by $Y_i(0) = R_i$ for all i. For each edge (i,j), define a matrix $Q^{(i,j)} \in \mathbb{R}^{n \times n}$ by

$$Q_\ell^{(i,j)} Y = \begin{cases} \frac{Y_i + Y_j}{2} & \text{if } \ell \in \{i, j\}, \\ Y_\ell & \text{otherwise.} \end{cases}$$

At each time t, choose an edge (i,j), and set $Y(k+1) = Q^{(i,j)}(Y(k))$. If the graph is connected and every edge is sampled infinitely often, then $\lim_{k \to \infty} Y(t) = \overline{Y}$, where $\overline{Y}_i = \overline{R}$. To see this, note that the operators $Q^{(i,j)}$ preserve the average value of the vector, hence

$$\frac{1}{n} \sum_{i=1}^{n} Y_i(k) = \overline{R}.$$

Further, for any k, either $Y(k+1) = Y(k)$ or $\|Y(k+1) - \overline{Y}\| < \|Y(k) - \overline{Y}\|$. Further, $\overline{Y}$ is the unique vector with average value $\overline{R}$ that is a fixed point for all operators $Q^{(i,j)}$. Hence, as long as the graph is connected and each edge is sampled infinitely often, $Y_i(k) \to \overline{R}$ as $k \to \infty$ and the components agree to the common average $\overline{R}$.

In the context of distributed optimization protocol, we will assume that each component i maintains a scalar value $Y_i(k)$ at time k representing an estimate of the total global reward. We will define a structure by which nodes communicate. In particular, for an ordered set of distinct edges $S = \big((i_i, j_1), \ldots, (i_{|S|}, j_{|S|})\big)$, we will define a set $\mathbb{W}_S \subset \mathbb{W}$. Let $\sigma(E)$ be the set of all possible ordered sets of disjoint edges S, including the empty set. We will assume that the sets $\{W_S | S \in \sigma(E)\}$ are disjoint and together form a partition of $\mathbb{W}$.

If $w(k) \in \mathbb{W}_S$, for some set S, we will assume that the components along the edges in S communicate in the order specified by S. Define

$$Q^S = Q^{(i_{|S|}, j_{|S|})} \cdots Q^{(i_1, j_1)},$$

where the terms in the product are taken over the order specified by S. Define $R(k) = (r_1(k), \ldots, r_n(k))$ as the vector of rewards occurring at time k. The update rule for the vector $Y(k)$ is given by

$$Y(k+1) = R(k+1) + \alpha Q^{S(k+1)} Y(k),$$

where $S(k+1)$ is the element of $\sigma(E)$ that contains $w(k+1)$. We will make the following assumption.

Assumption 6.1. *Define the set of edges $\hat{E}$ by*

$$\hat{E} = \{(i,j) | (i,j) \in S \text{ and } \mathbb{W}_S \neq \emptyset\}.$$

The graph $(V, \hat{E})$ is connected.

Since the process $(w(k), a(k))$ is aperiodic and has a single recurrent class (Assumption 1.1), this assumption guarantees that every edge on a connected subgraph is sampled infinitely often.

Policy parameters are updated at each component according to the rule:

$$\theta_i(k+1) = \theta_i(k) + \epsilon z_i^{\beta}(k)(1-\alpha)Y_i(k).$$

Note that, for this scheme, in relation to (4.1), we have

$$d_{ji}^{\alpha}(\ell, k) = n(1-\alpha)\alpha^{k-\ell} \left[\hat{Q}(\ell, k)\right]_{ij}, \tag{6.1}$$

where

$$\hat{Q}(\ell, k) = Q^{S(k-1)} \cdots Q^{S(\ell)},$$

Lemma 6.1. *The variables $d^{\alpha}_{ji}(\ell, k)$ defined by* (6.1) *satisfy Assumption 4.1.*

Proof. By definition, Assumption 4.1(1) is satisfied. Assumption 4.1(3) is also clearly satisfied.

Define the matrix $\mathcal{E}$ by $\mathcal{E}_{ij} = 1/n$ for all i, j. Then, Assumption 4.1(2) is equivalent to

$$\left\| \hat{Q}(\ell, k) - \mathcal{E} \right\| < c_\ell \gamma^{k-\ell}, \tag{6.2}$$

for a constant $\gamma \in (0, 1)$ and a random variable c_ℓ, such that the distribution of c_ℓ given $\mathcal{F}_\ell$ depends only on $(w(\ell), a(\ell))$, and with $\mathrm{E}[c_\ell | \mathcal{F}_\ell] \leq \bar{c}$ for a constant $\bar{c} < \infty$.

From Assumption 1.1 and Assumption 6.1, there must be some set of states $w_0, \ldots, w_{m-1}$ and corresponding edge sets $\bar{S}_0, \ldots, \bar{S}_{m-1}$, such that for each i, $w_i \in \mathbb{W}_{\bar{S}_i}$,

$$\bigcup_{i=0}^{m-1} \bar{S}_i = \hat{E},$$

and for some $\ell > 0$,

$$\Pr\left\{ w(\ell) = w_0, \ldots, w(\ell + m - 1) = w_{m-1} \right\} > 0.$$

Since this event occurs once with positive probability, it must occur infinitely often with probability 1. Define $N(k)$ to be the number of non-overlapping occurrences at or before time k, that is

$$N(k) = \sum_{\ell=0}^{k} \mathbf{1}_{\{w(\ell) \in \mathbb{W}_{\bar{S}_0}, \ldots, w(\ell+m-1) \in \mathbb{W}_{\bar{S}_{m-1}}\}}.$$

Define matrix $\overline{Q}$ and the set $\{(i_0, j_0), \ldots, (i_M, j_M)\}$ by

$$\overline{Q} = Q^{\bar{S}_m} \cdots Q^{\bar{S}_0} = \prod_{\ell=0}^{M} Q^{(i_\ell, j_\ell)}.$$

and let

$$\overline{\gamma} = \left\| \overline{Q} - \mathcal{E} \right\|.$$

We wish to show that $\overline{\gamma} < 1$. Assume otherwise, and let $\hat{x}$ be a vector such that $\|\hat{x}\| = 1$ and $\|(\overline{Q} - \mathcal{E})\hat{x}\| \geq 1$. Note that for every (i, j), $\mathcal{E} Q^{(i,j)} = Q^{(i,j)} \mathcal{E} = \mathcal{E}$, and $\mathcal{E}^2 = \mathcal{E}$. Hence,

$$\overline{Q} - \mathcal{E} = \prod_{\ell=0}^{M} \left(Q^{(i_\ell, j_\ell)} - \mathcal{E} \right).$$

Further, for any (i,j) and any vector x, either $Q^{(i,j)}x = x$ or $\|(Q^{(i,j)} - \mathcal{E})x\| < \|(I - \mathcal{E})x\|$. Since

$$\|(I-\mathcal{E})x\|^2 = x^T(I-\mathcal{E})x = x^T(I-\mathcal{E}^2)x = \|x\|^2 - \|\mathcal{E}x\|^2 \leq \|x\|^2,$$

we have

$$1 \leq \|(\overline{Q}-\mathcal{E})\hat{x}\| \leq \left\| \left(\prod_{\ell=0}^{M} \left(Q^{(i_\ell,j_\ell)} - \mathcal{E} \right) \right) \hat{x} \right\| \leq \prod_{\ell=0}^{M} \left\| Q^{(i_\ell,j_\ell)} - \mathcal{E} \right\| \leq 1.$$

Then, it follows that for every $(i,j) \in \hat{E}$, $Q^{(i,j)}\hat{x} = \hat{x}$. Since the set of edges $\hat{E}$ connects every node in the graph, if, for some pair of components p and q, $\hat{x}_p \neq \hat{x}_q$, we could construct a path of edges in $\hat{E}$ between p and q, and for some edge (i,j) along this path, $Q^{(i,j)}\hat{x} \neq \hat{x}$. Hence, the vector $\hat{x}$ must be constant. Then, $\|(\overline{Q}-\mathcal{E})\hat{x}\| = 0$. We have a contradiction, hence $\overline{\gamma} < 1$.

Set

$$t_\ell = \min\{k \geq 0 | N(k) = \ell\}.$$

Define $\overline{\Delta} = \mathrm{E}[t_{\ell+1} - t_\ell]$ (for $\ell \geq 1$) to be the expected time between non-overlapping observations of the communication pattern associated with $\overline{Q}$, and pick arbitrary $\epsilon \in (0,1)$ and $\delta \in (0, 1/\bar{\Delta})$. Define $\gamma = \bar{\gamma}^\delta \in (0,1)$, and note that $\overline{\gamma} < \gamma^{\bar{\Delta}} < 1$. Returning to (6.2), we have, for $\ell < k$,

$$\begin{aligned}
\left\| \hat{Q}(\ell,k) - \mathcal{E} \right\| &\leq \overline{\gamma}^{N(k-m)-N(\ell)} \mathbf{1}_{\{\ell < k-m\}} + \mathbf{1}_{\{\ell \geq k-m\}} \\
&\leq \gamma^{\bar{\Delta}(N(k-m)-N(\ell))} \mathbf{1}_{\{\ell < k-m\}} + \mathbf{1}_{\{\ell \geq k-m\}} \\
&< \Big(\gamma^{-(1-\epsilon)m} \gamma^{\bar{\Delta}(N(k-m)-N(\ell)) - (1-\epsilon)(k-m-\ell)} \mathbf{1}_{\{\ell < k-m\}} \\
&\qquad + \gamma^{-(1-\epsilon)m} \mathbf{1}_{\{\ell \geq k-m\}} \Big) \gamma^{(1-\epsilon)(k-\ell)} \\
&\leq c_\ell \gamma^{(1-\epsilon)(k-\ell)},
\end{aligned}$$

where

$$c_\ell = \gamma^{-(1-\epsilon)m} \left(1 + \sup_{\tau > 0} \gamma^{\bar{\Delta}(N(\ell+\tau)-N(\ell)) - (1-\epsilon)\tau} \right).$$

We wish to consider $\mathrm{E}[c_\ell | \mathcal{F}_\ell]$. Note that the distribution of c_ℓ given $\mathcal{F}_\ell$ depends only on $(w(\ell), a(\ell))$. It suffices consider the case where $\ell = 0$ over varying initial conditions $(w(0), a(0))$. Then, we have

$$
\begin{aligned}
&\mathrm{E}\left[\sup_\tau \gamma^{\bar{\Delta}N(\tau)-(1-\epsilon)\tau}\middle|\,\mathcal{F}_0\right] \\
&= \int_1^\infty \Pr\left\{\sup_\tau \gamma^{\bar{\Delta}N(\tau)-(1-\epsilon)\tau} > x\middle|\,\mathcal{F}_0\right\} dx \\
&= (-\log\gamma)\int_0^\infty \Pr\left\{\sup_\tau \gamma^{\bar{\Delta}N(\tau)-(1-\epsilon)\tau} > \gamma^{-u}\middle|\,\mathcal{F}_0\right\} \gamma^{-u} du \\
&= (-\log\gamma)\int_0^\infty \Pr\left\{\sup_\tau (1-\epsilon)\tau - \bar{\Delta}N(\tau) > u\middle|\,\mathcal{F}_0\right\} \gamma^{-u} du \\
&= (-\log\gamma)\int_0^\infty \left(1 - \Pr\left\{(1-\epsilon)\tau - \bar{\Delta}N(\tau) \le u, \forall\tau\,\middle|\,\mathcal{F}_0\right\}\right) \gamma^{-u} du
\end{aligned}
$$

Define

$$b_\ell = (1-\epsilon)t_\ell - \bar{\Delta}\ell,$$

and note that

$$\Pr\left\{\sup_\ell b_\ell \le u + (1+\epsilon) - \bar{\Delta}\middle|\,\mathcal{F}_0\right\} = \Pr\left\{(1-\epsilon)\tau - \bar{\Delta}N(\tau) \le u, \forall\tau\,\middle|\,\mathcal{F}_0\right\}.$$

Let $\Delta_\ell = (1-\epsilon)(t_{\ell+1} - t_\ell) - \bar{\Delta}$, so that $b_\ell = \sum_{s=0}^{\ell-1}\Delta_s$.

Since the process is generated by a finite state irreducible Markov chain, the tail of the interarrival times $t_{\ell+1} - t_\ell$ is bounded by a decaying exponential. Hence, the moment generating function $\mathrm{E}[e^{\eta\Delta_\ell}]$ of Δ_ℓ is finite for $\eta \in (-\infty, \overline{\eta})$ for some $\overline{\eta} > 0$. It follows that b_ℓ has a finite-valued moment generating function

$$\mathrm{E}[e^{\eta b_\ell}|\mathcal{F}_0] = \mathrm{E}[e^{\eta\Delta_0}|\mathcal{F}_0](\mathrm{E}[e^{\eta\Delta_1}])^{(\ell-1)},$$

for $\eta \in (-\infty, \overline{\eta})$. (Note that since the system is starting in an arbitrary initial state, Δ_0 has a different distribution than Δ_ℓ for $\ell > 0$.) By the Chernoff bound, for any $\eta \in (-\infty, \overline{\eta})$ and $x \ge 0$,

$$\Pr\left\{b_\ell \ge x|\,\mathcal{F}_0\right\} \le e^{-\eta x}\,\mathrm{E}[e^{\eta\Delta_0}|\mathcal{F}_0](\mathrm{E}[e^{\eta\Delta_1}])^{(\ell-1)} = e^{-\eta x + \rho_0(\beta) + (\ell-1)\rho_1(\eta)},$$

where $\rho_i(\eta) = \log\mathrm{E}[e^{\eta\Delta_i}]$. Since $\rho_1'(0) = \mathrm{E}[\Delta_1] = -\epsilon < 0$, there exist scalars $A > 0$, $\zeta > 0$ and $\kappa = -\rho(\zeta) > 0$ such that

$$\Pr\left\{b_\ell \ge x|\,\mathcal{F}_0\right\} \le Ae^{-\zeta x - \kappa\ell}.$$

Then,

$$
\begin{aligned}
1 - \Pr\left\{\sup_\ell b_\ell \le u(1+\epsilon) - \bar{\Delta}\right\} &\le \sum_{\ell=0}^{\infty} \Pr\left\{b_\ell > u - (1+\epsilon)\bar{\Delta}\right\} \\
&\le \sum_{\ell=0}^{\infty} A e^{-\zeta(u+(1+\epsilon)-\bar{\Delta})-\kappa k} \\
&= \frac{A}{1-e^{-\kappa}} e^{-\zeta(u+(1+\epsilon)-\bar{\Delta})}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\mathrm{E}\left[\sup_\tau \gamma^{\bar{\Delta}N(\tau)-(1-\epsilon)\tau} \middle| \mathcal{F}_0\right] \\
&= (-\log\gamma)\int_0^\infty \left(1 - \Pr\left\{(1-\epsilon)\tau - \bar{\Delta}N(\tau) \le u, \forall\tau \middle| \mathcal{F}_0\right\}\right)\gamma^{-u}du \\
&\le (-\log\gamma)\int_0^\infty \frac{A}{1-e^{-\kappa}} e^{-\zeta(u+(1+\epsilon)-\bar{\Delta})}\gamma^{-u}du.
\end{aligned}
$$

The final term is finite if $\gamma > e^{-\zeta}$. Note, however, by choosing δ sufficiently small, γ can be made arbitrarily close to 1. Hence, for such a choice of γ, $\mathrm{E}[c_0|\mathcal{F}_0]$ is finite. $\square$

7 Convergence Analysis

We will first introduce tools from the theory of stochastic approximation. Using these tools, we will be able to establish the convergence of the two algorithms presented earlier.

7.1 Stochastic Approximation

Stochastic approximation provides an iterative method to solve equations of the form

$$\bar{g}(\theta) = 0$$

for some continuous function $\bar{g}(\theta)$. In our instance, if we set $\bar{g}(\theta) = \nabla_\theta \lambda(\theta)$, stochastic approximation will allow us to find policy parameters which are local optima of the expected average reward function.

In particular, consider the iterative scheme

$$\theta(k+1) = \theta(k) + \epsilon g(\theta(k), \xi(k)). \tag{7.1}$$

Here, $g(\theta(k), \xi(k))$ is an estimate of $\bar{g}(\theta(k))$ at time k, and $\xi(k)$ is a process that captures the underlying state and whatever additional noise memory is required to compute the estimate. In our framework, we will require that $\xi(k)$ has a Markov structure: given $\theta(k)$, the distribution of $\xi(k+1)$ depends only on $\xi(k)$. In other words,

$$\Pr\left(\xi(k+1) \in \cdot | \mathcal{F}_k\right) = \mathcal{P}\left(\xi(k), \cdot | \theta(k)\right), \tag{7.2}$$

for some transition function $\mathcal{P}$.

We have not yet defined the relationship between the estimators $g(\theta, \xi)$ and the function $\bar{g}(\theta)$. We will require that, when θ is held fixed, the values $g(\theta, \xi(k))$ locally average to $\bar{g}(\theta)$. In order to make this notion precise, note that for a fixed value of θ, the transition function $\mathcal{P}(\cdot, \cdot|\theta)$ defines a Markov chain we shall call the fixed-θ chain and denote by $\xi_\theta(k)$. The local averaging condition requires that

$$\lim_{K \to \infty} \frac{1}{K} \operatorname{E}\left[\sum_{k=0}^{K-1} g(\theta, \xi_\theta(k))\right] = \bar{g}(\theta), \tag{7.3}$$

for each initial condition $\xi_\theta(0)$.

Consider the ordinary differential equation

$$\dot{\bar{\theta}}(t) = \bar{g}(\bar{\theta}(t)). \tag{7.4}$$

Define $\mathcal{L}$ to be the set of limit points of (7.4) over all initial conditions. Let $\theta^\epsilon(k)$ be the sequence of parameters resulting from (7.1) with a particular fixed ϵ. Finally, define a continuous-time interpolation $\bar{\theta}^\epsilon(t)$ if $\theta^\epsilon(k)$ by setting $\bar{\theta}^\epsilon(t) = \theta^\epsilon(k)$ if $t \in [k\epsilon, k\epsilon + \epsilon)$. In the following lemma, we will establish conditions for the weak convergence of $\bar{\theta}^\epsilon(t)$ to a solution $\bar{\theta}(t)$ of the ODE (7.4) as $\epsilon \to 0$, such that the fraction of the time interval $[0, T]$ that $\theta^\epsilon(t)$ spends in a small neighborhood of $\mathcal{L}$ will go to 1 in probability as $\epsilon \to 0$ and $T \to \infty$.

Note that when $\bar{g}(\theta) = \nabla_\theta \lambda(\theta)$, the function $\lambda(\theta)$ is a Lyapunov function for the ODE. Then, the set of limit points L is the set of stationary points θ for which

$$\nabla_\theta \lambda(\theta) = 0.$$

Hence, the limit points are local optima of $\lambda(\theta)$.

Lemma 7.1. *Assume the following conditions:*

1. *The iterates $\{\theta^\epsilon(k) | k, \epsilon\}$ are bounded.*
2. *There exists an $\mathcal{F}_t$-measurable process $\xi(t) \in I \subset \Xi$, where I is a compact set in a complete separable metric space Ξ, and a transition function $\mathcal{P}(\cdot, \cdot|\theta)$ such that the Markov condition* (7.2) *holds.*

3. *$\mathcal{P}(\xi, \cdot|\theta)$ is weakly continuous in (θ, ξ), that is, for every bounded and continuous real-valued function F on $\Re^S$, the value of the integral*

$$\int F(\tilde{\xi})\mathcal{P}(\xi, d\tilde{\xi}|\theta)$$

is continuous in (θ, ξ).

4. *The set of invariant measures under transition functions $\mathcal{P}(\xi, \cdot|\theta)$ is tight over all θ.*

5. *The estimate function $g(\theta, \xi)$ is continuous, bounded, and measurable, and satisfies the local averaging condition* (7.3) *for a fixed-θ chain.*

Then, for any sequence of processes $\{\bar{\theta}^\epsilon(t)|\epsilon\to 0\}$ there exists a subsequence that weakly converges to $\bar{\theta}(t)$ as $\epsilon\to 0$, where $\bar{\theta}(t)$ is a solution to the ODE (7.4). *Further, for $\delta > 0$, define $N_\delta(\mathcal{L})$ to be a neighborhood of radius δ around the limit set $\mathcal{L}$. The fraction of time that $\hat{\theta}^\epsilon(t)$ spends in $N_\delta(\mathcal{L})$ over the time interval $[0, T]$ goes to 1 in probability as $\epsilon\to 0$ and $T\to\infty$.*

Proof. The result follows directly from Theorem 8.4.3 in [2]. □

7.2 Convergence of the Distributed Algorithm

We wish to prove convergence of the stochastic approximation scheme corresponding to our distributed optimization algorithm:

$$\theta_i^\epsilon(k+1) = \theta_i^\epsilon(k) + \epsilon z_i^\beta(k)(1-\alpha)Y_i(k). \tag{7.5}$$

Theorem 7.1. *Assume that the set of iterates $\{\theta^\epsilon(k)|k, \epsilon\}$ from* (7.5) *are bounded. Then, the conclusions of Lemma 7.1 hold.*

Proof. We will use the framework provided by Lemma 7.1. Define

$$\xi(k) = (w(k), a(k), z_1^\beta(k), \ldots, z_n^\beta(k), Y(k)),$$

$\Xi = \mathbb{X} \times \mathbb{A} \times \mathbb{R}^{N+n}$. To see that $\xi(t)$ takes values in a compact subset of Ξ, it suffices to prove that $z_i^\beta(k)$ and $Y(k)$ are bounded. Yet,

$$\left\|z_i^\beta(k)\right\| = \left\|\sum_{\ell=0}^{k} \beta^{k-\ell}\nabla_i(\ell)\right\| \le L\sum_{\ell=0}^{k}\beta^{k-\ell} \le \frac{L}{1-\beta},$$

$$\|Y(k)\| = \left\|\sum_{\ell=0}^{k}\alpha^{k-\ell}\hat{Q}(\ell,k)R(\ell)\right| \le \frac{\hat{R}}{1-\alpha},$$

where

$$\bar{R} = \max_{w \in \mathbb{W}, a \in \mathbb{A}} \left| \sum_{i=1}^{n} r_i(w,a)^2 \right|^{1/2}.$$

Further, since $(w(k), a(k))$ form a Markov chain and from (3.2) and the definition of $Y(k)$, clearly $\xi(t)$ is an $\mathcal{F}_k$-measurable Markov chain. The fact that the associated transition function is weakly continuous follows from the smoothness conditions on $\pi_\theta(a|x)$ provided by Assumption 2.1. Define the function

$$g(\theta, \xi) = \left(g_1(\theta, \xi), \ldots, g_n(\theta, \xi)\right),$$

where

$$g_i(\theta, \xi) = (1 - \alpha) z_i(t) y_i(t).$$

Boundedness of $g(\theta, \xi)$ is clear, further

$$\chi_i(t) = g_i(\theta(t), \xi(t)).$$

Finally, Theorem 5.1 provides the appropriate averaging condition for the fixed-θ chain. □

References

[1] J. Baxter and P. L. Bartlett. Infinite-Horizon Gradient-Based Policy Search. *Journal of Artificial Intelligence Research*, 15:319–350, 2001.

[2] H. J. Kushner and G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, New York, NY, 1997.

[3] P. Marbach and J.N. Tsitsiklis. Simulation–Based Optimization of Markov Reward Processes. *IEEE Transactions on Automatic Control*, 46(2):191–209, 2001.