Errata

“Asynchronous Stochastic Approximation and Q-Learning”
J. N. Tsitsiklis,


The proof of Lemma 9 is incorrect as written. A corrected version, essentially the same as the one given in D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-dynamic Programming*, Athena Scientific, 1996, Proposition 5.6, is as follows.

The definition of $F_{iu}^\pi$ in p. 200 should be

$$F_{iu}^\pi(Q) = E[c_{iu}] + \sum_{j \neq 1} p_{ij}(u)Q_{j,\pi(j)}, \quad i \neq 1, u \in U(i).$$

Then, consider a Markov chain with states $(i, u)$ and with the following dynamics: from any state $(i, u)$, we move to state $(j, \pi(j))$, with probability $p_{ij}(u)$; in particular, subsequent to the first transition, we are always at a state of the form $(i, \pi(i))$ and the first component of the state evolves according to $\pi$. Let us identify all states of the form $(1, u)$, with a single (absorbing) state. Because $\pi$ was assumed proper for the original problem, it follows that the system with states $(i, u)$ also evolves according to a proper policy. The transition probability matrix for this chain, after deleting the row and column associated with the absorbing state, has a maximal eigenvalue strictly less than one. By the Perron-Frobenius theorem, there exists a positive vector $w$ with components $w_{i,u}$ and some $\gamma \in [0,1)$ such that

$$\sum_{j \neq 1} p_{ij}(u)w_{j,\pi(j)} \leq \gamma w_{i,u}, \quad \forall i \neq 1.$$ 

Therefore, for any vectors $Q$ and $Q'$, we have

$$\frac{|F_{iu}^\pi(Q) - F_{iu}^\pi(Q')|}{w_{i,u}} \leq \frac{1}{w_{i,u}} \sum_{j \neq 1} p_{ij}(u)w_{j,\pi(j)} \frac{|Q_{j,\pi(j)} - Q'_{j,\pi(j)}|}{w_{j,\pi(j)}} \leq \gamma \max_{j \neq 1, u \in U(j)} \frac{|Q_{ju} - Q'_{ju}|}{w_{j,u}}.$$ 

The rest of the argument remains as given in the paper.