A APPENDIX

Lemma 1. *The vector field $f : \mathbb{R}^{K+D} \to \mathbb{R}^K$ such that*

$$f(v,u) = \mathbb{E}\Big[\big(R + \gamma \max_{a' \in \mathcal{A}} Q_{v,u}(X', a') - \xi Q_{v,u}(X, A)\big)\phi_u(X, A)\Big] - \epsilon v$$

is Lipschitz-continuous.

Proof. We must show that

$$\|f(v,u) - f(w,z)\| \leq L_f \|v - w\| + L_f \|u - z\| .$$

For a fixed transition tuple $(x, a, r, x') \in \mathcal{X} \times \mathcal{A} \times I_R \times \mathcal{X}$, we define $f_{x,a,r,x'} : \mathbb{R}^{K+D} \to \mathbb{R}^K$ such that

$$f_{x,a,r,x'}(v,u) = \big(r + \gamma \max_{a' \in \mathcal{A}} Q_{v,u}(x', a') - \xi Q_{v,u}(x, a)\big)\phi_u(x, a) + \epsilon v.$$

We show that $f_{x,a,r,x'}$ is Lipschitz-continuous with with a constant L_f that does not depend on (x, a, r, x'). Consequently, $f = \mathbb{E}[f_{x,a,r,x'}]$ is also Lipschitz-continuous with the same constant.

We assume (x, a, r, x') is fixed and we use the notation:

- $A(v,u) = r + \gamma \max_{a' \in \mathcal{A}} Q_{v,u}(x', a')$;
- $B(v,u) = \xi Q_{v,u}(x, a)$;
- $C(u) = \phi_u(x, a)$;
- $D(v) = \epsilon v$.

Then we have that $f_{x,a,r,x'}(v,u) = A(v,u)C(u) - B(v,u)C(u)$. We move on using triangle inequalities.

$$\begin{aligned}
\|f_{x,a,r,x'}(v,u) - f_{x,a,r,x'}(w,z)\| = \\
&= \|A(v,u)C(u) - B(v,u)C(u) - A(w,z)C(z) + B(w,z)C(z) + D(v) - D(w)\| \\
&\leq \|A(v,u)C(u) - A(w,z)C(z)\| + \|B(v,u)C(u) - B(w,z)C(z)\| + \\
&\quad + \|D(v) - D(w)\| \\
&\leq \big\|A(v,u)\big(C(u) - C(z)\big)\big\| + \big\|\big(A(v,u) - A(w,z)\big)C(z)\big\| + \\
&\quad + \big\|B(v,u)\big(C(u) - C(z)\big)\big\| + \big\|\big(B(v,u) - B(w,z)\big)C(z)\big\| \\
&\qquad + \|D(v) - D(w)\| .
\end{aligned}$$

We observe that

$$\|A(v,u)\| \leq r_{max} + \gamma \max_{a' \in \mathcal{A}} \|\phi_u(x', a')\| \, \|Proj(v)\| \leq r_{max} + \gamma\rho$$

using Cauchy-Schwartz inequality, Assumption (ii) and the projection of v. Also,

$$\begin{aligned}
\|C(u) - C(z)\| &\leq \|\phi_u(x, a) - \phi_z(x, a)\| \\
&\leq L_\phi \|u - z\|
\end{aligned}$$

from Assumption (ii). We also have that

$$\begin{aligned}
\|A(v,u) - A(w,z)\| &\leq \gamma \|Proj(v) - Proj(w)\| + \|Proj(w)\| \, \|u - z\| \\
&\leq \gamma \|v - w\| + \rho \|u - z\|
\end{aligned}$$

for the same reasons. From assumption (iii),

$$\|C(u)\| = 1.$$

Now,

$$\|B(v,u)\| \leq \xi \|\phi_u(x, a)\| \, \|Proj(v)\| \leq \xi\rho$$

and

$$\begin{aligned}\|B(v,u)-B(w,z)\| &\leq \left\|\xi\big(\phi_u(x,a)-\phi_z(x,a)\big)Proj(u)\right\| + \|\xi\phi_z(x,a)(Proj(v)-Proj(z))\| \\ &\leq \xi\rho L_\phi \|u-z\| + \xi\|v-w\|.\end{aligned}$$

Finally, we can see that

$$\|D(v)-D(w)\| = \epsilon\|v-w\|.$$

Putting everything together with the help of Cauchy-Schwartz inequalities, the conclusion follows. □

Lemma 2. *The sequence of random vectors $\{M_t\}_{t\geq 0}$ such that*

$$\begin{aligned}M_{t+1} = \big(r_t + \gamma \max_{a'\in\mathcal{A}} Q_{v_t,u_t}(x'_t,a'_t) - \xi Q_{v_t,u_t}(x_t,a_t)\big)\phi_{u_t}(x_t,a_t) - \\ - \mathbb{E}\Big[\big(R+\gamma\max_{a'\in\mathcal{A}} Q_{v,u}(X',a') - \xi Q_{v,u}(X,A)\big)\phi_u(X,A)\Big]\end{aligned}$$

is a martingale difference sequence verifying

$$\mathbb{E}\big[\,\|M_{t+1}\|^2 \mid \mathcal{F}_t\big] \leq c_M\big(1+\|v_t\|^2+\|u_t\|^2\,\big).$$

Proof. For $\{M_t\}_{t\geq 0}$ to be a martingale difference sequence, we must have that

1. $\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] = 0, \quad \forall t > 0;$
2. $\mathbb{E}[\|M_t\|] < \infty, \quad \forall t > 0.$

From Assumption (i), we directly conclude property 1:

$$\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] = \mathbb{E}[M_{t+1}] = 0.$$

To establish property 2, we observe that every term appearing on the definition of M_t is bounded by some constant. The same observation is sufficient to conclude the second moment is also bounded by some constant. □

Lemma 3. *For each $u \in \mathbb{R}^D$, the o.d.e.*

$$\dot{v}_t = f(v_t,u)$$

has a unique and globally asymptotically stable equilibrium $v^*(u)$, where $v^* : \mathbb{R}^D \to \mathbb{R}^K$ is Lipschitz-continuous.

Proof. We start by establishing the existence and uniqueness of an equilibrium, $v^*(u)$, for each $u \in \mathbb{R}^D$, by making use of Banach's fixed point theorem. We then show that v^* is Lipschitz-continuous. Finally, we show that $v^*(u)$ is globally asymptotically stable using a Lyapunov argument.

Any solution to the o.d.e. must verify

$$\dot{v}_t = f(v_t,u) = 0.$$

Equivalently, and ignoring for a moment the projection of $Proj(v)$ of v in Q, any solution is time-invariant and we can, therefore, drop the dependency on t and, writing in the form of a fixed-point equation, we must have that

$$v = \frac{1}{\xi}\Sigma_u^{-1}\mathbb{E}\Big[\big(R+\gamma\max_{a'\in\mathcal{A}} Q_{v,u}(X',a')\big)\phi_u(X,A)\Big] - \frac{\epsilon}{\xi}\Sigma_u^{-1}v.$$

We used assumption (iv) to invert the matrix Σ_u. Let us refer to the right-hand side as $T : \mathbb{R}^K \to \mathbb{R}^K$. As R^K is a Banach space, contractiveness of T will allow us to conclude the existence and uniqueness of solution to the fixed point equation above. We observe that

$$\begin{aligned}\|T(v)-T(w)\| &\leq \left\|\frac{1}{\xi}\Sigma_u^{-1}\gamma\mathbb{E}\Big[\big(\max_{a'\in\mathcal{A}}Q_{v,u}(X',a')-\max_{a'\in\mathcal{A}}Q_{w,u}(X',A')\big)\phi_u(X,A)\Big]\right\|+\\ &\quad+\left\|\frac{\epsilon}{\xi}\Sigma_u^{-1}(v-w)\right\|\\ &\leq \frac{\gamma}{\xi\sigma}\mathbb{E}\Big[\max_{a'\in\mathcal{A}}|\phi_u(X',a')\cdot v-\phi_u(X',a')\cdot w|\,\|\phi_u(X,A)\|\Big]+\\ &\quad+\frac{\epsilon}{\xi\sigma}\|v-w\|\\ &\leq \frac{\gamma+\epsilon}{\xi\sigma}\|v-w\|\end{aligned}$$

From assumption (iv), $\frac{\gamma+\epsilon}{\xi\sigma}<1$ and contractiveness holds. Therefore, there exists a unique solution $v^*\in\mathbb{R}^K$ for each $u\in\mathbb{R}^D$. Importantly, we can also obtain that $\|v^*(u)\|\leq\frac{1}{\xi\sigma-\epsilon}(r_{max}+\gamma\rho)<\rho$. Therefore, the solution of the o.d.e. is inside the ball B_ρ.

Now, we show that the solution $v^*:\mathbb{R}^D\to\mathbb{R}^K$ obtained is Lipshitz-continuous on $u\in\mathbb{R}^K$. The proof is very similar to the one of conctrativeness of T. For that, we write

$$\begin{aligned}\|v^*(u)-v^*(w)\| &\leq \frac{\gamma}{\xi\sigma}\mathbb{E}\Big[\max_{a'\in\mathcal{A}}|\phi_u(X',a')\cdot v^*(u)-\phi_w(X',a')\cdot v^*(w)|\,\|\phi_u(X,A)\|\Big]\\ &\leq \frac{\gamma\rho}{\xi\sigma}L_\phi\|u-w\|.\end{aligned}$$

Finally, having established uniqueness and existence of a Lipschitz-solution of the o.d.e., $v^*(u)$, we prove it is globally asymptotically stable. We consider the Lyapunov function $l_u:\mathbb{R}^K\to\mathbb{R}$ such that $l_u(v)=\frac{1}{2}\|v-v^*(u)\|^2$. We have that $l_u(v)=0$ if and only if $v=v^*(u)$. We also have that $l_u(v)>0$ if and only if $v\neq v^*(u)$. To establish globally asymptotic stability, it remains only show that $\dot{l}_u(v)<0$ whenever $v\neq v^*(u)$ and $\dot{l}_u(v)=0$ otherwise. We start by writing

$$\begin{aligned}\dot{l}_u(v) &= \nabla_v l_u(v)\cdot\dot{v}\\ &=\big(v-v^*(u)\big)\cdot f(v,u)\\ &=\big(v-v^*(u)\big)\cdot\mathbb{E}\Big[\big(R+\gamma\max_{a'\in\mathcal{A}}Q_{v,u}(X',a')-\xi Q_{v,u}(X,A)\big)\phi_u(X,A)+\epsilon v\Big]\\ &=\big(v-v^*(u)\big)\cdot\mathbb{E}\Big[\big(R+\gamma\max_{a'\in\mathcal{A}}Q_{v,u}(X',a')\big)\phi_u(X,A)+\epsilon v\Big]-\xi\big(v-v^*(u)\big)\Sigma_u v.\end{aligned}$$

Now, we subtract the quantity $\mathbb{E}\Big[\big(R+\gamma\max_{a'\in\mathcal{A}}Q_{v^*(u),u}(X',a')\big)\phi_u(X,A)+\epsilon v^*(u)\Big]-\xi\Sigma_u v^*(u)$, which we know equals 0, multiplied by $(v-v^*(u))$. We can rearrange the resulting expression and obtain

$$\begin{aligned}&\big(v-v^*(u)\big)\cdot\mathbb{E}\Big[\big(R+\gamma\max_{a'\in\mathcal{A}}Q_{v,u}(X',a')-\gamma\max_{a'\in\mathcal{A}}Q_{v^*(u),u}(X',a')\big)\phi_u(X,A)+\epsilon\big(v-v^*(u)\big)\Big]-\\ &\quad-\xi\big(v-v^*(u)\big)\Sigma_u\big(v-v^*(u)\big).\end{aligned}$$

Since $\xi>0$ is sufficiently large, we can conclude that $\dot{l}_u(v)<0$ if $\big(v-v^*(u)\big)\Sigma_u\big(v-v^*(u)\big)>0$. Such is the case since we have positive-definiteness of Σ_u from assumption (iv). This concludes the result. □

Lemma 4. *For every $u\in\mathbb{R}^K$, the sequence of vector fields $\{h_{c,u}\}_{c\geq1}$ such that $h_{c,u}:\mathbb{R}^K\to\mathbb{R}^K$ and $h_{c,u}(v)=\frac{f(cv,u)}{c}$, for some continuous $h_{\infty,u}$, verifies*

$$h_{c,u}\to h_{\infty,u}.$$

uniformly on compacts. Additionally, the o.d.e.

$$\dot{v}_t=h_{\infty,u}(v_t)$$

has the origin has its unique and globally asymptotically stable equilibrium.

Proof. We can expand the definition and observe

$$h_{c,u}(v) = \frac{\mathbb{E}\Big[\big(R + \gamma \max_{a' \in \mathcal{A}} Q_{cv,u}(X', a') - \xi Q_{cv,u}(X, A)\big)\phi_u(X, A)\Big] - \epsilon cv}{c}.$$

We recall that Q projects v back into B_ρ once $v > \rho$. Therefore, as $c \to \infty$,

$$h_{c,u}(v) \to -\epsilon v$$

uniformly on compacts. With $h_{\infty,u}(v) = -\epsilon v$, we have that $\dot{v}_t = h_{\infty,u}(v_t)$ has the origin as unique and globally asymptotically stable equilibrium. □

Lemma 5. *Let $V^* = \{\big((v^*(u), u\big), u \in \mathbb{R}^D\}$. If $\sup_{t \geq 0} \|u_t\| < \infty$, then $(v_t, u_t) \to V^* \quad w.p.1.$*

Proof. Having established Lemmas 1 to 5, we can use Theorem 2 from Borkar (2008, Chapter 6). □

Lemma 6. *The finite composition of Lipschitz-continuous and Lipschitz-smooth function is Lipschitz-continuous and Lipschitz-smooth.*

Proof. Let $f, g : X \to Y$ such that f and g are Lipschitz-continuous with constants L_f and L_g and Lipschitz-smooth with constants $L_{\dot{f}}$ and $L_{\dot{g}}$ respectively. Notice that the derivatives of f and g are therefore bounded by C_f nd C_g.

We show that $f \circ g$ is Lipschitz-continuous.

$$\begin{aligned}
\big\|(f \circ g)(x) - (f \circ g)(z)\big\| &= \big\|f\big(g(x)\big) - f\big(g(y)\big)\big\| \\
&\leq L_f \|g(x) - g(y)\| \\
&\leq L_f L_g \|x - y\|.
\end{aligned}$$

We can also show $f \circ g$ is Lipschitz-smooth. We present the proof for the one-dimensional case.

$$\begin{aligned}
\Big\|\big(f \circ g\big)'(x) - \big(f \circ g\big)'(z)\Big\| &= \big\|f'\big(g(x)\big)g'(x) - f'\big(g(y)\big)g'(y)\big\| \\
&= \big\|f'\big(g(x)\big)g'(x) - f'\big(g(y)\big)g'(y) + f'\big(g(x)\big)g'(y) - f'\big(g(x)\big)g'(y)\big\| \\
&\leq \big\|f'\big(g(x)\big)g'(x) - f'\big(g(x)\big)g'(y)\big\| + \big\|f'\big(g(x)\big)g'(y) - f'\big(g(y)\big)g'(y)\big\| \\
&\leq \big\|f'\big(g(x)\big)\big\| \|g'(x) - g'(y)\| + \|g'(y)\| \big\|f'\big(g(x)\big) - f'\big(g(y)\big)\big\| \\
&\leq C_f \dot{L}_g \|x - y\| + C_g L_{\dot{f}} \|g(x) - g(y)\| \\
&\leq C_f \dot{L}_g \|x - y\| + C_g L_{\dot{f}} L_g \|x - y\|
\end{aligned}$$

□