On Differentiability of Average Cost in Parameterized Markov Chains

Vijay Konda
John N. Tsitsiklis

August 30, 2002

1 Overview

The purpose of this appendix is to prove Theorem 4.6 in [5] and establish various facts used in verifying some of the assumptions of Theorem A.1 therein. We use the same notation and assumptions as in [5], and to simplify the analysis, we assume that the parameter θ is scalar. The setting used here is similar (but not identical) to the one in [4]. The major difference is that the setting of [4] allows only RSPs of the form

$$U_k = f(X_k, W_k),$$

where the parameter θ enters only through the distribution of W_k which does not depend on x. In contrast, our setting allows for any RSPs of the form

$$U_k = f_\theta(X_k, W_k),$$

as long as the distribution $\mu_\theta(u|x)$ of $f_\theta(x, U)$ satisfies the required assumptions.

The differentiability results presented in this report are based on the following key lemma (see [3, Corollary 2.2.1]).

Lemma 1.1. *Consider a parameterized family of random variables $\{F_\theta(\omega); \theta \in \mathbb{R}\}$ on a Polish space Ω. Suppose the following hold:*

1. *For each θ_0, there exists some $d > 1$ and $K_d > 0$ such that,*

$$\mathbf{E}\left[|F_\theta - F_{\theta_0}|^d\right] < K_d |\theta - \theta_0|^d,$$

for all θ sufficiently close to θ_0.

2. *There is a family of random variables $\{f_\theta\}$ such that*

$$\frac{F_{\theta+h} - F_\theta}{h} \xrightarrow{\mathbf{P}} f_\theta, \quad \text{as } h \to 0, \quad \forall \theta,$$

where $\xrightarrow{\mathbf{P}}$ denotes convergence in probability.

Then the map $\theta \mapsto \mathbf{E}[F_\theta]$ is differentiable with

$$\frac{d}{d\theta}\mathbf{E}[F_\theta] = \mathbf{E}[f_\theta].$$

Our proof of the differentiability results involves the following steps.

1. Obtain a representation of the steady-state expected cost and of solutions to the Poisson equation in terms of suitably defined expectations $\mathbf{E}_{\theta,x}$.

2. Using likelihood ratios, represent the above expectations using only the expectations $\mathbf{E}_{\theta_0,x}$ corresponding to some fixed θ_0. This new representation has the advantage that the probability law w.r.t. which expectations are taken does not depend on the parameter θ. Only the random variables inside the expectation depend on the parameter θ, making our key result Lemma 1.1 applicable to this representation.

3. Verify the hypotheses of Lemma 1.1.

The next section carries out these steps for expectations over finite horizon. In Section 3, we recall the results on the regenerative representations for the average cost and solutions to the Poisson equation. In Section 4, we introduce the likelihood ratio representations and show that they satisfy the assumptions of Lemma 1.1. In Sections 5 and 6, we use all the previous results to prove differentiability of expectations over infinite horizon and verify some of the assumptions of Theorem A.1 in [5].

2 Finite Horizon

In this section, we prove the differentiability of expectations of the form $\mathbf{E}_{\theta,x}[W]$ where W is a random variable that depends only on a finite number of state-decision pairs (X_k, U_k). For each θ and k, let

$$\Lambda_{\theta,k} = \prod_{l=0}^{k} \frac{\mu_\theta(U_l|X_l)}{\mu_{\theta_0}(U_l|X_l)}.$$

It is easy to see that $\Lambda_{\theta,k}$ is the Radon-Nikodym derivative of the distribution of $(X_l, U_l)_{l=0}^{k}$ under policy θ with respect to the distribution under policy θ_0. We will use the following basic lemma to derive some bounds on $\Lambda_{\theta,k}$.

LEMMA 2.1. *If $W_0, \dots, W_{k-1}$ are positive random variables satisfying*

$$\mathbf{E}\left[|W_l - 1|^d\right] \leq \epsilon^d C(d), \quad \forall d \geq 1, \quad l = 0, \dots, k-1,$$

for some monotonic function $C(\cdot) \geq 1$, and if $\epsilon < 1/k2^k$, then

$$\mathbf{E}\left[\left|\prod_{l=0}^{k-1} W_l - 1\right|^d\right] \leq \epsilon^d (2k)^d C(kd),$$

$$\mathbf{E}\left[\prod_{l=0}^{k-1} W_l^d\right] \leq 2^d C(kd),$$

for any $d \geq 1$.

Proof. For any $d \geq 1$ and $z \geq 0$, we have

$$z^d \leq 2^{d-1}\left(1 + |z-1|^d\right), \tag{1}$$

we have

$$\begin{aligned}\mathbf{E}\left[W_l^d\right] &\leq 2^{d-1}(1 + \epsilon^d C(d)) \\ &\leq 2^d C(d), \quad l = 0, \dots, k-1.\end{aligned}$$

To prove the first part, note that

$$|z_0 \cdots z_{k-1} - 1| \leq |z_0 - 1| + z_0|z_1 - 1| + \cdots + z_0 \cdots z_{k-2}|z_{k-1} - 1|$$

and therefore

$$|z_0 \cdots z_{k-1} - 1|^d \leq k^{d-1}\left[|z_0 - 1|^d + z_0^d|z_1 - 1|^d + \cdots + z_0^d \cdots z_{k-2}^d|z_{k-1} - 1|^d\right].$$

We now use Holder's inequality to see that

$$\begin{aligned}\mathbf{E}\left[\left|\prod_{l=0}^{k-1} W_l - 1\right|^d\right] &\leq k^{d-1}\sum_{l=0}^{k-1}\mathbf{E}\left[\prod_{j=0}^{l-1} W_j^d\,|W_l - 1|^d\right] \\ &\leq k^{d-1}\sum_{l=0}^{k-1}\prod_{j=0}^{l-1}\mathbf{E}\left[W_j^{(l+1)d}\right]^{1/(l+1)}\mathbf{E}\left[|W_l - 1|^{(l+1)d}\right]^{1/(l+1)} \\ &\leq k^{d-1}\sum_{l=0}^{k-1} 2^{ld}\epsilon^d C((l+1)d) \\ &\leq k^d\epsilon^d 2^{kd} C(kd).\end{aligned}$$

To prove the second part, use (1) and the fact that $\epsilon < 1/k2^k$ to simplify the bound. □

The following are two immediate consequences of the above lemma. In the following lemma, note that the constant K_d might depend on k. This dependence can be safely ignored as the next lemma is intended to verify the first assumption of Lemma 1.1 for a fixed k.

Lemma 2.2.

1. *For every $d \geq 1$, there exists $K_d > 0$ such that*

$$\mathbf{E}_{\theta_0,x}\left[|\Lambda_{\theta,k} - 1|^d\right] \leq K_d|\theta - \theta_0|^d L(x), \quad \forall\theta, x.$$

2. *For every $d \notin (1, -1)$, there exists $K_d > 0$ such that*

$$\mathbf{E}_{\theta_0,x}\left[|\Lambda_{\theta,k}|^d\right] \leq K_d L(x), \quad \forall x,$$

and for all θ sufficiently close to θ_0.

Proof. Use the mean-value theorem to see that

$$\left| \frac{\mu_\theta(u|x)}{\mu_{\theta_0}(u|x)} - 1 \right| \le |\theta - \theta_0| \sup_{\bar{\theta},\theta_0} \left| \frac{\nabla \mu_{\bar{\theta}}(u|x)}{\mu_{\theta_0}(u|x)} \right|. \tag{2}$$

Since $\nabla \mu_\theta(u|x)/\mu_{\theta_0}(u|x)$ belongs to $\mathcal{D}$, we have

$$\mathbf{E}_{\theta,x} \left[\sup_{\bar{\theta},\theta_0} \frac{\nabla \mu_{\bar{\theta}}(X_k, U_k)}{\mu_{\theta_0}(X_k, U_k)} \right] \le K_d L(x)$$

for some constant K_d. Part 1, and Part 2 for $d \ge 1$, are implied by Lemma 2.1. To prove Part 2 for $d \le -1$, note that the inequality (2) holds even if θ and θ_0 are interchanged. □

The above lemma verifies the first assumption of Lemma 1.1 for expectations of random variables that are functions of state-decision pairs up to time k. The next lemma verifies the second assumption and derives a formula for the gradient.

Lemma 2.3. *Consider a family of random variables $\{W_\theta\}$ such that the map $\theta \mapsto W_\theta(\omega)$ is differentiable for each ω, W_θ is a function of the state-decision pairs up to time k, and*

$$\mathbf{E}_{\theta,x}[|W_\theta|^d] < \infty, \qquad \mathbf{E}_{\theta,x} \left[\sup_{\bar{\theta}} |\nabla W_{\bar{\theta}}|^d \right] < \infty \quad \forall \theta, x,$$

for some $d > 1$. Then, the map $\theta \mapsto \mathbf{E}_{\theta,x}[W_\theta]$ is differentiable with

$$\nabla \mathbf{E}_{\theta,x} \left[W_\theta \right] = \mathbf{E}_{\theta,x} \left[\nabla W_\theta \right] + \mathbf{E}_{\theta,x} \left[\Lambda'_{\theta,k} W_\theta \right],$$

where

$$\Lambda'_{\theta,k} = \sum_{l=0}^{k} \psi_\theta(X_l, U_l) = \sum_{l=0}^{k} \frac{\nabla \mu_\theta(X_l, U_l)}{\mu_\theta(X_l, U_l)}.$$

Proof. Note that

$$\mathbf{E}_{\theta,x} \left[W_\theta \right] = \mathbf{E}_{\theta_0,x} \left[\Lambda_{\theta,k} W_\theta \right].$$

Using the bounds of the previous lemma, the fact that d-moments of W_θ and $\sup_\theta \nabla W_\theta$ are finite, and the mean-value theorem, verify the first assumption of Lemma 1.1. Using the fact that $\mu_\theta(u|x)$ and W_θ are differentiable in θ, the second assumption of Lemma 1.1 follows easily. The gradient formula now follows from interchanging the order of differentiation and expectation. □

The following is an immediate consequence of the previous result. It uses the formula in the previous result to derive bounds on the rate at which the gradient

$$\nabla \mathbf{E}_{\theta,x} \left[f_\theta(X_k, U_k) \right]$$

grows with k.

Corollary 2.4. *Consider a family of functions $\{f_\theta(x,u)\}$ in $\mathcal{D}$ such that the family $\{\nabla f_\theta(x,u)\}$ is also in $\mathcal{D}$. Then, the map $\theta \mapsto \mathbf{E}_{\theta,x}\left[f_\theta(X_k,U_k)\right]$ is differentiable for each x, and there exists some $K > 0$ such that*

$$|\nabla \mathbf{E}_{\theta,x}\left[f_\theta(X_k,U_k)\right]| \le (k+1)KL(x), \quad \forall x.$$

Furthermore, the families of functions

$$\left\{\mathbf{E}_{\theta,x}\left[f_\theta(X_k,U_k)\right]\right\}, \qquad \left\{\nabla \mathbf{E}_{\theta,x}\left[f_\theta(X_k,U_k)\right]\right\}$$

belong to $\mathcal{D}$.

Before we move on, we would like to point out that the above result holds for expectations of the form $\mathbf{E}_{\theta,x}[\cdot|U_0 = u]$ as well, if we redefine $\Lambda'_{\theta,k}$ as

$$\Lambda'_{\theta,0} = 0, \qquad \Lambda'_{\theta,k} = \sum_{l=1}^{k} \psi_\theta(X_l,U_l), \quad \forall k > 0.$$

3 Splitting

In this section, we recall the splitting technique of Athreya and Ney [2], and Nummelin [7] to obtain a regenerative representation for the average cost function $\bar{\alpha}(\theta)$ and the solutions Q_θ of the Poisson equation. Let δ, N, $\mathbb{X}_0$ and ϑ be as in Assumption 4.2 of [5]. Consider the $\{0,1\}$-valued process (B_k) constructed as follows:

1. If k is not divisible by N, then $B_k = 0$.
2. If k is divisible by N, then

$$\begin{aligned}\mathbf{P}(B_k = 1|X_l,U_l,l=0,1,\dots,B_l,l\neq k) &= \mathbf{P}(B_k = 1|X_k,X_{k+N})\\ &= f_\theta^{(1)}(X_k,X_{k+N})\end{aligned}$$

where

$$f_\theta^{(1)}(x,y) = \frac{\delta}{2} \times I_{\mathbb{X}_0}(x)\frac{\vartheta(dy)}{\mathbf{P}_{\theta,x}(X_N \in dy)}.$$

(Assumption 4.2(a) ensures existence of $f_\theta^{(1)}$.)

Let

$$\begin{aligned}f_\theta^{(0)}(x,y) &= 1 - f_\theta^{(1)}(x,y), \quad \forall \theta,x,y,\\ Q_{\theta,x}(dy) &= \frac{\mathbf{P}_{\theta,x}(X_N \in dy) - \frac{\delta}{2} \times I_{\mathbb{X}_0}(x)\vartheta(dy)}{1 - \frac{\delta}{2} \times I_{\mathbb{X}_0}(x)}.\end{aligned}$$

It is easy to see that for each θ, x,

$$f_\theta^{(0)}(x,\cdot) = \left(1 - \frac{\delta}{2} \times I_{\mathbb{X}_0}(x)\right) \times \frac{Q_{\theta,x}(dy)}{\mathbf{P}_{\theta,x}(X_N \in \cdot dy)}$$

where the existence of the Radon-Nikodym derivative is guaranteed by the fact that

$$Q_{\theta,x}(dy) \quad \leq \quad \frac{2}{2-\delta}\mathbf{P}_{\theta,x}(X_N \in dy).$$

With this construction it is not difficult to see that the process (X_{kN}, B_{kN}) is a Markov chain for which $\mathbb{X} \times \{1\}$ is an atom. Consider the first time this atom is hit, *i.e.,*

$$\tau = \min\{k | B_k = 1\}.$$

The time τ is well-defined and finite w.p.1. as the set $\mathbb{X}_0$ is hit infinitely often from any initial state (cf. Assumption 4.2 in [5]) and the probability that B_k is 1 given that $X_k = x$ is $\delta/2$ whenever $x \in \mathbb{X}_0$. Using standard results [6, Theorems 10.0.1 and 10.2.2] on Markov chains with atoms, the average cost $\bar{\alpha}(\theta)$ can be written as

$$\bar{\alpha}(\theta) = \frac{\mathbf{E}_{\theta,\vartheta}\left[\sum_{k=0}^{(\tau/N)-1} \bar{c}_{\theta,N}(X_{kN})\right]}{\mathbf{E}_{\theta,\vartheta}\left[\tau\right]},$$

where

$$\bar{c}_{\theta,N}(x) = \sum_{k=0}^{N-1} \mathbf{E}_{\theta,x}\left[c(X_k, U_k)\right].$$

Furthermore, the functions $V_\theta(x)$ given by

$$V_\theta(x) = \mathbf{E}_{\theta,x}\left[\sum_{k=0}^{(\tau/N)-1} \left(\bar{c}_{\theta,N}(X_{kN}) - N\bar{\alpha}(\theta)\right)\right] \tag{3}$$

satisfy the Poisson equation

$$V_\theta(x) = \mathbf{E}_{\theta,x}\left[c(x, U_0) + V(X_1)\right]$$

for the Markov chain $\{X_k\}$. To see this, note that $V_\theta(x)$ is a solution to the Poisson equation for the cost function $\bar{c}_{\theta,N}$ and the Markov chain $\{X_{kN}\}$ (see [6, p.441]). Furthermore, if $\hat{V}_\theta$ is a solution to the Poisson equation for the cost function $\bar{c}_{\theta,1}$ and the Markov chain $\{X_k\}$, then $\hat{V}_\theta$ is a solution to the Poisson equation for the cost function $\bar{c}_{\theta,N}$ and the Markov chain $\{X_{kN}\}$. Since $\hat{V}_\theta$ and V_θ are two solutions to the Poisson equation for a positive Harris chain $\{X_{kN}\}$, they must differ by a constant (cf. Proposition 17.4.1 of [6]) and therefore, V_θ is a solution to the Poisson equation for the cost function $\bar{c}_{\theta,1}$ and the Markov chain $\{X_k\}$.

4 Likelihood Ratio

For each θ, k, let

$$\tilde{\Lambda}_{\theta,k} = \Lambda_{\theta,kN}\hat{\Lambda}_{\theta,k}$$

where

$$\hat{\Lambda}_{\theta,k} = \prod_{l=0}^{k-1} \frac{f_\theta^{(0)}(X_{lN}, X_{lN+N})}{f_{\theta_0}^{(0)}(X_{lN}, X_{lN+N})},$$

and 0/0 is interpreted as 1. It is easy to see that $\tilde{\Lambda}_{\theta,k}$ is the Radon-Nikodym derivative of the joint distribution of $(X_l, U_l; 0 \leq l \leq kN, B_l; 0 \leq l \leq (k-1)N)$ under $\mathbf{P}_\theta$ with respect to that under $\mathbf{P}_{\theta_0}$ on the set of outcomes for which $\tau > kN$. Therefore for any W that is a function of only these random variables, we have

$$\mathbf{E}_{\theta,x}\left[WI\{\tau > kN\}\right] = \mathbf{E}_{\theta_0,x}\left[WI\{\tau > kN\}\tilde{\Lambda}_{\theta,k}\right].$$

To show that the map

$$\theta \mapsto \mathbf{E}_{\theta,x}\left[WI\{\tau > kN\}\right]$$

is differentiable, we need to verify the hypothesis of Lemma 1.1 for the likelihood ratio $\tilde{\Lambda}_{\theta,k}$. We now derive several bounds to verify the first assumption of Lemma 1.1.

Lemma 4.1.

1. *For each $d \geq 1$, there exists $K_d > 0$ such that*

$$\mathbf{E}_{\theta_0,x}\left[\left|\frac{f_\theta^{(i)}(x, X_N)}{f_{\theta_0}^{(i)}(x, X_N)} - 1\right|^d\right] \leq K_d|\theta - \theta_0|^d L(x) \quad \forall\theta, x, i = 0, 1.$$

2. *For each $d \geq 1$, there exists $K_d > 0$ such that*

$$\mathbf{E}_{\theta_0,x}\left[\left|\frac{f_\theta^{(i)}(x, X_N)}{f_{\theta_0}^{(i)}(x, X_N)}\right|^d\right] \leq K_d L(x) \quad \forall x, i = 0, 1,$$

   and for all θ sufficiently close to θ_0.

3. *For each $d \geq 1$, there exists $K_d > 0$ such that*

$$\mathbf{E}_{\theta_0,x}\left[\left|\hat{\Lambda}_{\theta,k} - 1\right|^d\right] \leq K_d|\theta - \theta_0|^d L(x) \quad \forall\theta, x.$$

4. *For each $d \geq 1$, there exists $K_d > 0$ such that*

$$\mathbf{E}_{\theta_0,x}\left[\left|\hat{\Lambda}_{\theta,k}\right|^d\right] \leq K_d L(x) \quad \forall x,$$

   and for all θ sufficiently close to θ_0.

Proof. To prove the first part, note that for $x \notin \mathbb{X}_0$,

$$\frac{f_\theta^{(1)}(x,y)}{f_{\theta_0}^{(1)}(x,y)} = 1,$$

and for $x \in \mathbb{X}_0$,

$$\begin{aligned}\frac{f_\theta^{(1)}(x,y)}{f_{\theta_0}^{(1)}(x,y)} &= \frac{\mathbf{P}_{\theta_0,x}(X_N \in dy)}{\mathbf{P}_{\theta,x}(X_N \in dy)} \\ &= \left\{\mathbf{E}_{\theta_0,x}\left[\Lambda_{\theta,N}|\, X_N = y\right]\right\}^{-1}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathbf{E}_{\theta_0,x}\left[\left|\frac{f_\theta^{(1)}(x,X_N)}{f_{\theta_0}^{(1)}(x,X_N)} - 1\right|^d\right] &\le \mathbf{E}_{\theta_0,x}\left[\left|\mathbf{E}_{\theta_0,x}\left[\Lambda_{\theta,N}|\, X_N\right]^{-1} - 1\right|^d\right] \\ &\le \sqrt{\mathbf{E}_{\theta_0,x}\left[\mathbf{E}_{\theta_0,x}\left[\Lambda_{\theta,N}|\, X_N\right]^{-2d}\right]\mathbf{E}_{\theta_0,x}\left[\left|\mathbf{E}_{\theta_0,x}\left[\Lambda_{\theta,N} - 1|\, X_N\right]\right|^{2d}\right]} \\ &\le \sqrt{\mathbf{E}_{\theta_0,x}\left[\Lambda_{\theta,N}^{-2d}\right]\mathbf{E}_{\theta_0,x}\left[\left|\Lambda_{\theta,N} - 1\right|^{2d}\right]},\end{aligned}$$

where the last two inequalities follow from Holder's and Jensen's inequalities respectively. For $i = 0$, note that

$$\begin{aligned}\left|\frac{f_\theta^{(0)}(x,y)}{f_{\theta_0}^{(0)}(x,y)} - 1\right| &= \left|\frac{f_\theta^{(1)}(x,y)}{f_{\theta_0}^{(1)}(x,y)} - 1\right|\left(\frac{f_{\theta_0}^{(1)}(x,y)}{f_{\theta_0}^{(0)}(x,y)}\right) \\ &\le 2\left|\frac{f_\theta^{(1)}(x,y)}{f_{\theta_0}^{(1)}(x,y)} - 1\right|\end{aligned}$$

since $f_\theta^{(1)} \le 1/2$ for all θ. The proof of the second part is similar. The last two parts follow from the first two and Lemma 2.1. □

The next lemma proves that the first of the two hypotheses for Lemma 1.1 hold for $\tilde{\Lambda}$.

LEMMA 4.2.

1. *For each $d \ge 1$, there exists $K_d > 0$ such that*

$$\mathbf{E}_{\theta_0,x}\left[\left|\tilde{\Lambda}_{\theta,k} - 1\right|^d\right] \le K_d|\theta - \theta_0|^d L(x),$$

for θ sufficiently close to θ_0.

2. *For each $d \ge 1$, there exists $K_d > 0$ such that*

$$\mathbf{E}_{\theta_0,x}\left[\left|\tilde{\Lambda}_{\theta,k}\right|^d\right] \le K_d L(x).$$

Proof. Follows from Lemmas 2.2, 4.1 and 2.1. □

We are now ready to state the following result on differentiability of finite horizon expectations.

LEMMA 4.3. *For any family of functions $\{f_\theta(x,u)\}$ in $\mathcal{D}$ such that the family $\{\nabla f_\theta(x,u)\}$ is also in $\mathcal{D}$, the map $\theta \mapsto \mathbf{E}_{\theta,x}\left[\bar{f}_\theta(X_{kN})I\{\tau > kN\}\right]$, where*

$$\bar{f}_\theta(x) = \sum_{k=0}^{N-1} \mathbf{E}_{\theta,x}\left[f_\theta(X_k, U_k)\right],$$

is differentiable for each x, with

$$\left|\nabla \mathbf{E}_{\theta,x}\left[\bar{f}_\theta(X_{kN})I\{\tau > kN\}\right]\right| \leq (k+1)\rho_0^k K L(x),$$

for some $\rho_0 < 1$.

Proof. Since

$$\mathbf{E}_{\theta,x}\left[\bar{f}_\theta(X_{kN})I\{\tau > kN\}\right] = \mathbf{E}_{\theta_0,x}\left[\tilde{\Lambda}_{\theta,k}\bar{f}_\theta(X_{kN})I\{\tau > kN\}\right],$$

we can use Lemma 2.1 to prove its differentiability and to calculate the derivative. To verify the first assumption of Lemma 2.1, use Holder's inequality to see that

$$\begin{aligned} &\mathbf{E}_{\theta_0,x}\left[\left|\bar{f}_\theta(X_{kN})I\{\tau > kN\}\tilde{\Lambda}_{\theta,k} - \bar{f}_\theta(X_{kN})I\{\tau > kN\}\tilde{\Lambda}_{\theta_0,k}\right|^d\right] \\ = \quad &\mathbf{E}_{\theta_0,x}\left[\left|\bar{f}_\theta(X_{kN})\right|^d\left|\tilde{\Lambda}_{\theta,k} - 1\right|^d\right] \\ \leq \quad &\mathbf{E}_{\theta_0,x}\left[\left|\bar{f}_\theta(X_{kN})\right|^{2d}\right]^{1/2}\mathbf{E}_{\theta_0,x}\left[\left|\tilde{\Lambda}_{\theta,k} - 1\right|^{2d}\right]^{1/2}. \end{aligned}$$

The first assumption now follows from the fact that $f_\theta(x,u)$ (and therefore $\bar{f}_\theta(x)$) belongs to $\mathcal{D}$, and the first part of the previous lemma. To verify the second assumption of Lemma 1.1, use Assumption 4.4 of [5] to conclude that the map

$$\theta \mapsto \Lambda_{\theta,k}(\omega)$$

is continuously differentiable for all ω, and therefore it is enough to prove that

$$\frac{\hat{\Lambda}_{\theta_0+h,k} - 1}{h}$$

converges in probability. This in turn is verified if the functions

$$f_\theta^{(i)}(x,y)/f_{\theta_0}^{(i)}(x,y), i = 0, 1,$$

can be shown to be differentiable w.r.t. θ for all x, y. Since

$$f_{\theta_0}^{(1)}(x,y)/f_\theta^{(1)}(x,y)$$

is equal to

$$\mathbf{E}_{\theta_0,x}\left[\Lambda_{\theta,N}\right| X_N = y] ,$$

for all $x \notin \mathbb{X}_0$ and is equal to 1 otherwise, use part 3 of Assumption 4.4 in [5] and the mean-value theorem to see that

$$\frac{\Lambda_{\theta_0+h,N} - 1}{h}$$

is bounded above by an integrable random variable. Therefore, we invoke dominated convergence theorem for conditional expectations to show that there exists a version of

$$f_\theta^{(1)}(x,y)/f_{\theta_0}^{(1)}(x,y)$$

that is differentiable at θ_0 for all x, y. Furthermore,

$$\frac{\tilde{\Lambda}_{\theta_0+h} - 1}{h}$$

converges in probability to $\tilde{\Lambda}'_{\theta_0,k}$ where

$$\tilde{\Lambda}'_{\theta,k} \;=\; \sum_{l=0}^{kN} \psi_\theta(X_k, U_k) - \sum_{j=0}^{k-1} I_{\mathbb{X}_0}(X_{jN}) \mathbf{E}_{\theta,x}\left[\sum_{l=jN}^{jN+N-1} \psi_\theta(X_l, U_l) \right| X_{jN}, X_{jN+N} \Bigg] .$$

Therefore, we have

$$\begin{aligned} \nabla \mathbf{E}_{\theta,x}\left[\bar{f}_\theta(X_{kN}) I\{\tau > kN\}\right] &= \mathbf{E}_{\theta,x}\left[\tilde{\Lambda}'_{\theta,k} \bar{f}_\theta(X_{kN}) I\{\tau > kN\}\right] \\ &\quad + \mathbf{E}_{\theta,x}\left[\tilde{\Lambda}_{\theta,k} \nabla \bar{f}_\theta(X_{kN}) I\{\tau > kN\}\right] \end{aligned}$$

To derive the bound on the derivative, use Holder's inequality to obtain

$$\begin{aligned} \left|\nabla \mathbf{E}_{\theta,x}\left[\bar{f}_\theta(X_{kN}) I\{\tau > k\}\right]\right| &\leq \mathbf{E}_{\theta,x}\left[\left|\tilde{\Lambda}'_{\theta,k}\right|^3\right]^{1/3} \mathbf{E}_{\theta,x}\left[\left|\bar{f}_\theta(X_{kN})\right|^3\right]^{1/3} \\ &\quad \cdot \mathbf{P}_{\theta,x}(\tau > kN)^{1/3} \\ &\quad + \mathbf{E}_{\theta,x}\left[\left|\nabla \bar{f}_\theta(X_{kN})\right|^3\right]^{1/3} \cdot \mathbf{P}_{\theta,x}(\tau > kN)^{1/3}. \end{aligned}$$

To calculate

$$\mathbf{E}_{\theta,x}\left[\left|\tilde{\Lambda}'_{\theta,k}\right|^3\right]^{1/3},$$

note that $\tilde{\Lambda}'_{\theta,k}$ is a sum of $2kN+1$ terms. Since $\psi_\theta(x,u)$ belongs to $\mathcal{D}$, the 3-norm of each of these terms is bounded by $KL(x)^{1/3}$ for some K. Similarly, since $f_\theta(x,u)$ and $\nabla f_\theta(x,u)$ (therefore $\bar{f}_\theta(x)$ and $\nabla \bar{f}_\theta(x)$) belong to $\mathcal{D}$, the terms

$$\mathbf{E}_{\theta,x}\left[\left|\bar{f}_\theta(X_{kN})\right|^3\right]^{1/3} \text{ and } \mathbf{E}_{\theta,x}\left[\left|\nabla \bar{f}_\theta(X_{kN})\right|^3\right]^{1/3}$$

are also bounded by $KL(x)^{1/3}$ for some K. Finally, due to uniform geometric ergodicity, we have

$$\mathbf{P}_{\theta,x}(\tau > kN) \leq K\rho_0^k L(x),$$

for some $\rho_0 < 1$. Combining all these terms it is easy to establish the stated bound on the derivative. □

5 Infinite Horizon

The above lemma concerns the differentiability of expectation w.r.t. finite dimensional marginals of $\mathbf{P}_{\theta,x}$. We use the following result from advanced calculus (e.g., see Apostol [1]) to extend the above result to the infinite horizon case.

Theorem 5.1. *Assume that each f_k is a real-valued function defined on a neighborhood Θ of θ_0 such that the derivative $f_k'(\theta)$ exists for each θ in Θ. Assume that*

1. *$\sum f_k(\theta_0)$ converges,*
2. *there exists a $g(\theta)$ such that $\sum \nabla f_k(\theta) = g(\theta)$ uniformly on Θ.*

Then:

1. *There exists a function $f(\theta)$ such that $\sum f_k(\theta) = f(\theta)$ uniformly on Θ.*
2. *If $\theta \in \Theta$, the derivative $\nabla f(\theta)$ exists and equals $\sum \nabla f_k(\theta)$.*

By combining the previous two results we have the following theorem. Recall that η_θ is the steady state expectation of the state-decision process $\{(X_k, U_k)\}$.

Theorem 5.2. *For any family of functions $\{f_\theta(x,u)\}$ in $\mathcal{D}$ such that the family $\{\nabla f_\theta(x,u)\}$ is also in $\mathcal{D}$, the map*

$$\theta \mapsto \mathbf{E}_{\theta,x}\left[\sum_{k=0}^{(\tau/N)-1} \bar{f}_\theta(X_{kN})\right]$$

is differentiable for each x, and the families of functions

$$\left\{\mathbf{E}_{\theta,x}\left[\sum_{k=0}^{(\tau/N)-1} \bar{f}_\theta(X_{kN})\right]\right\}, \left\{\nabla\mathbf{E}_{\theta,x}\left[\sum_{k=0}^{(\tau/N)-1} \bar{f}_\theta(X_{kN})\right]\right\},$$

belong to $\mathcal{D}$.

Proof. Note that

$$\mathbf{E}_{\theta,x}\left[\sum_{k=0}^{(\tau/N)-1} \bar{f}_\theta(X_{kN})\right] = \sum_{k=0}^{\infty} \mathbf{E}_{\theta,x}\left[\bar{f}_\theta(X_{kN}) I\{\tau > kN\}\right].$$

The differentiability now follows from the previous two lemmas. To show that the functions

$$\left\{\mathbf{E}_{\theta,x}\left[\sum_{k=0}^{(\tau/N)-1} \bar{f}_\theta(X_{kN})\right]\right\}, \left\{\nabla\mathbf{E}_{\theta,x}\left[\sum_{k=0}^{(\tau/N)-1} \bar{f}_\theta(X_{kN})\right]\right\}$$

are in $\mathcal{D}$, consider

$$\begin{aligned}
&\left|\mathbf{E}_{\theta,x}\left[\sup_{\bar\theta}\sum_{k=0}^{(\tau/N)-1}\bar f_{\bar\theta}(X_{kN})\right]\right|^d \\
&\quad\le \mathbf{E}_{\theta,x}\left[\left|\sum_{k=0}^{(\tau/N)-1}\sup_{\bar\theta}\bar f_{\bar\theta}(X_{kN})\right|^d\right] \\
&\quad\le \mathbf{E}_{\theta,x}\left[(\tau/N)^{d-1}\sum_{k=0}^{(\tau/N)-1}\left|\sup_{\bar\theta}\bar f_{\bar\theta}(X_{kN})\right|^d\right] \\
&\quad\le \sum_{k=0}^{\infty}\mathbf{E}_{\theta,x}\left[(\tau/N)^{d-1}\left|\sup_{\bar\theta}\bar f_{\bar\theta}(X_{kN})\right|^d I\{\tau > kN\}\right] \\
&\quad\le \sum_{k=0}^{\infty}\mathbf{E}_{\theta,x}\left[(\tau/N)^{3(d-1)}\right]^{1/3}\mathbf{E}_{\theta,x}\left[\left|\sup_{\bar\theta}\bar f_{\bar\theta}(X_{kN})\right|^{3d}\right]^{1/3} \\
&\qquad\quad \times\mathbf{P}_{\theta,x}(\tau > kN)^{1/3}.
\end{aligned}$$

It is now easy to see that the right hand side is bounded by $K_d L(x)$ for some $K_d > 0$. □

The following corollaries are immediate consequences of this result.

COROLLARY 5.3. *For any family of functions $\{f_\theta(x,u)\}$ in $\mathcal{D}$ such that the family $\{\nabla f_\theta(x,u)\}$ is also in $\mathcal{D}$, the map*

$$\theta \to \eta_\theta(f_\theta)$$

is bounded and differentiable with bounded derivatives.

COROLLARY 5.4. *(Lemma 5.1 in [5]) If $\mathbf{X}_0$ is singleton (i.e., $\mathbb{X}_0$ is an atom and therefore splitting in unnecessary) and τ is the first hitting time of $\mathbb{X}_0$ then the functions defined as*

$$\begin{aligned}
Q_\theta(x,u) &= \mathbf{E}_{\theta,x}\left[\sum_{k=0}^{\tau-1} c(X_k,U_k)\middle| U_0 = u\right], \\
T_\theta(x,u) &= \mathbf{E}_{\theta,x}\left[\tau\middle| U_0 = u\right],
\end{aligned}$$

belong to $\mathcal{D}$. Furthermore, for each (x,u), the maps $\theta \mapsto Q_\theta(x,u)$, $\theta \mapsto T_\theta(x,u)$ are differentiable, and the derivatives $\nabla Q_\theta(x,u)$, $\nabla T_\theta(x,u)$ also belong to $\mathcal{D}$.

COROLLARY 5.5. *The average expected cost function $\bar\alpha(\theta)$ is bounded and differentiable with bounded derivatives. Furthermore, the solutions $V_\theta(x)$ to the Poisson equation given by (3) belong to $\mathcal{D}$ and the map*

$$\theta \mapsto V_\theta(x)$$

is differentiable for each x and the derivatives $\nabla V_\theta(x)$ also belong to $\mathcal{D}$. Furthermore, the family of functions

$$Q_\theta = c - \bar\alpha(\theta)\underline{1} + P_\theta V_\theta$$

belongs to $\mathcal{D}$, with the map $\theta \to Q_\theta(x,u)$ being differentiable for all (x,u) and $\nabla Q_\theta(x,u)$ belonging to $\mathcal{D}$.

This corollary is used in [5] to establish Theorem 4.6 in [5], which is again proved here for completeness.

THEOREM 5.6. *For **any** solution Q_θ to the Poisson equation we have*

$$\nabla\bar{\alpha}(\theta) = \langle \psi_\theta, Q_\theta \rangle_\theta.$$

Furthermore, $\nabla\bar{\alpha}(\theta)$ has bounded derivatives.

Proof. Note that the Q_θ defined in the previous corollary satisfies the Poisson equation:

$$Q_\theta(x,u) = c(x,u) - \bar{\alpha}(\theta) + \mathbf{E}_{\theta,x}\left[Q_\theta(X_1,U_1) \middle| U_0 = u\right].$$

Since Q_θ belongs to $\mathcal{D}$, Q_θ is differentiable in θ, and ∇Q_θ also belongs to $\mathcal{D}$, we can differentiate both sides of the above equation to obtain

$$\begin{aligned}\nabla Q_\theta(x,u) &= -\nabla\bar{\alpha}(\theta) + \mathbf{E}_{\theta,x}\left[\nabla Q_\theta(X_1,U_1) \middle| U_0 = u\right] \\ &\quad + \mathbf{E}_{\theta,x}\left[\psi_\theta(X_1,U_1) Q_\theta(X_1,U_1) \middle| U_0 = u\right].\end{aligned}$$

Taking expectation with respect to the steady state distribution of (X_k, U_k) we obtain the desired formula with Q_θ defined as in the previous corollary. To see that Q_θ can be replaced by any other solution $\hat{Q}_\theta$ to the Poisson equation, note that $Q_\theta - \hat{Q}_\theta$ is some function $C(\theta)$ (Proposition 17.4.1 of [6]) and that

$$\mathbf{E}_{\theta,x}\left[\psi_\theta(x, U_0)\right] = 0.$$

To prove that $\bar{\alpha}_\theta$ is twice differentiable with bounded derivatives, note that $\nabla\bar{\alpha}_\theta = \eta_\theta(\psi_\theta Q_\theta)$ and apply Corollary 5.3. □

The above result leads to the following theorem on differentiability of a common representation for solutions to Poisson equation.

THEOREM 5.7. *For any family of functions $\{f_\theta(x,u)\}$ in $\mathcal{D}$ such that the family $\{\nabla f_\theta(x,u)\}$ is also in $\mathcal{D}$, the map*

$$\theta \mapsto \sum_{k=0}^{\infty} \mathbf{E}_{\theta,x}\left[f_\theta(X_k,U_k) - \eta_\theta(f_\theta)\right]$$

is differentiable for each x, and the families of functions

$$\left\{\sum_{k=0}^{\infty} \mathbf{E}_{\theta,x}\left[f_\theta(X_k,U_k) - \eta_\theta(f_\theta)\right]\right\}, \left\{\nabla\left[\sum_{k=0}^{\infty} \mathbf{E}_{\theta,x}\left[f_\theta(X_k,U_k) - \eta_\theta(f_\theta)\right]\right]\right\},$$

belong to $\mathcal{D}$.

Proof. To simplify the proof, assume that $f_\theta = c$. Then the results on differentiability for the finite horizon case, imply that the function

$$\theta \mapsto \mathbf{E}_{\theta,x}\left[c(X_k,U_k) - \bar{\alpha}(\theta)\right]$$

is differentiable with

$$\begin{aligned}
&\left|\nabla \mathbf{E}_{\theta,x}\left[c(X_k,U_k) - \bar{\alpha}(\theta)\right]\right| \\
&= \left|\sum_{l=0}^{k} \mathbf{E}_{\theta,x}\left[\psi_\theta(X_l,U_l)\left(c(X_k,U_k) - \bar{\alpha}(\theta)\right)\right] - \nabla\bar{\alpha}(\theta)\right| \\
&\le \sum_{l=0}^{[k/2]} \mathbf{E}_{\theta,x}\left[\psi_\theta(X_l,U_l)\mathbf{E}_{\theta,X_{l+1}}\left[\left|c(X_k,U_k) - \bar{\alpha}(\theta)\right|\right]\right] \\
&\quad + \sum_{l=[k/2]+1}^{k} \mathbf{E}_{\theta,x}\left[\left|\psi_\theta(X_l,U_l)\left(c(X_k,U_k) - \bar{\alpha}(\theta)\right)\right.\right. \\
&\qquad\qquad\qquad \left.\left. - \left\langle \psi_\theta, P_\theta^{k-l}(c - \bar{\alpha}(\theta)\underline{1})\right\rangle_\theta\right|\right] \\
&\quad + \sum_{l=[k/2]}^{\infty} \left\langle \psi_\theta, P_\theta^{l}(c - \bar{\alpha}(\theta)\underline{1})\right\rangle,
\end{aligned}$$

where $[k/2]$ represents the "floor" of $k/2$. Using geometric ergodicity, the fact that $V^{1/d}$ also satisfies the geometric Foster Lyapunov criterion with the decay factor $\rho^{1/d}$, and Schwartz inequality, we can see that each of these terms can be bounded by $K\rho_0^k L(x)$ for some $\rho_0 < 1$. The result now follows from an easy application of Theorem 5.1. □

6 Verification of Assumption A.3 parts (d) and (e)

In this section, we verify parts (d) and (e) of Assumption A.3 of [5] for the two cases of TD(1) and of TD(λ), $\lambda < 1$ separately. Since $\phi_\theta, Q_\theta, T_\theta$ belong to $\mathcal{D}$, part (d) in both cases follows from Corollary 5.3. Therefore, we will verify only part (e). We start with the TD(1) case.

6.1 TD(1)

We will now verify part (e) for $\hat{h}_\theta(\cdot)$. Note that

$$\begin{aligned}
\hat{h}_\theta(y) &= \mathbf{E}_{\theta,\bar{x}}\left[\sum_{k=0}^{\tau-1}(h_\theta(Y_k) - \bar{h}(\theta)) \;\middle|\; Y_0 = y\right], \\
&= \sum_{k=0}^{\infty} \mathbf{E}_{\theta,\bar{x}}\left[(h_\theta(Y_k) - \bar{h}(\theta))I\{\tau > k\} \;\middle|\; Y_0 = y\right].
\end{aligned}$$

Since $h_\theta(Y_k)$ is a function of the state-decision pairs up to time k and satisfies the assumptions of Lemma 2.3, a formula can be derived for the derivative of the expectation

$$\mathbf{E}_{\theta,\bar{x}}\left[(h_\theta(Y_k) - \bar{h}(\theta))I\{\tau > k\} \;\middle|\; Y_0 = y\right].$$

Using this formula, a bound on this derivative of the form

$$K(k+1)^d \rho_0^k \tilde{L}(x,u)z \quad , d \geq 1, K > 0, \rho_0 < 1, \tilde{L} \in \mathcal{D}.$$

can be obtained. This bound and Lemma 5.1 can be used to conclude that $\hat{h}_\theta$ is differentiable with respect to θ and the derivative is bounded by $K\tilde{L}(x,u)z$ for some other $K > 0$. It is now easy to verify that

$$\mathbf{E}_{\theta,x}\left[\hat{h}_\theta(Y_1)\Big|Y_0 = y\right]$$

is differentiable with the derivative bounded appropriately. The verification of part (e) for $\hat{G}_\theta$ is similar.

6.2 TD(λ), $\lambda < 1$

Since the verification of part (e) for $\hat{h}_\theta$ is simpler and the verification for $\hat{G}_\theta$ is similar, we will consider $\hat{h}_\theta$ only. Note that the first component of the vector $\hat{h}_\theta$ does not depend on z and is equal to

$$L\sum_{k=0}^{\infty}\mathbf{E}_{\theta,x}\left[(c(X_k,U_k) - \bar{\alpha}(\theta))\Big|U_0 = u\right].$$

Therefore, part (e) for this component follows from Lemma 5.7. For the second component, consider the sum

$$\begin{aligned}
&\sum_{k=0}^{\infty}\mathbf{E}_{\theta,x}\big[c(X_k,U_k)Z_k - \bar{h}_1(\theta) - \bar{\alpha}(\theta)\bar{Z}(\theta)\big|Y_0 = y\big] \\
&\quad = z\sum_{k=0}^{\infty}\lambda^k\mathbf{E}_{\theta,x}\left[c(X_k,U_k) - \langle P_\theta^l c, \phi_\theta\rangle_\theta\big|Y_0 = y\right] \\
&\qquad + \sum_{k=0}^{\infty}\sum_{l=0}^{k-1}\lambda^l\mathbf{E}_{\theta,x}\big[c(X_k,U_k)\phi_\theta(X_{k-l},U_{k-l}) - \langle P_\theta^l c, \phi_\theta\rangle_\theta\big|Y_0 = y\big] \\
&\qquad\quad + \sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty}\lambda^l\langle P_\theta^l c, \phi_\theta\rangle_\theta \\
&\quad = z\sum_{k=0}^{\infty}\lambda^k\mathbf{E}_{\theta,x}\left[c(X_k,U_k) - \langle P_\theta^l c, \phi_\theta\rangle_\theta\big|Y_0 = y\right] \\
&\qquad + \sum_{l=0}^{\infty}\lambda^l\sum_{k=l+1}^{\infty}\mathbf{E}_{\theta,x}\big[c(X_k,U_k))\phi_\theta(X_{k-l},U_{k-l}) - \langle P_\theta^l c - \bar{\alpha}(\theta)\underline{1}, \phi_\theta\rangle_\theta\big|Y_0 = y\big] \\
&\qquad\quad + \sum_{l=0}^{\infty}\lambda^l\sum_{k=0}^{l-1}\langle P_\theta^l c, \phi_\theta\rangle_\theta.
\end{aligned}$$

It is now easy to see that the second sums of the first two terms are differentiable with the derivatives uniformly bounded in l as the second sum of the second term is an expectation of the solution to a Poisson equation for the Markov chain $(X_{k-l}, U_{k-l}, X_k, U_k)$ (cf. Lemma 5.7). The second sum of the third term is also differentiable with the bound on the derivative linear in k. It now follows from Theorem 5.1 that $\hat{h}_\theta(y)$ is differentiable with the derivative bounded above by $K\tilde{L}(x,u)z$ for some $\tilde{L} \in \mathcal{D}$.

References

[1] T. M. Apostol. *Mathematical Analysis: A Modern Approach to Advanced Calculus.* Addison-Wesley Pub Co, 2 edition, January 1974.

[2] K. B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.*, 245:493–501, 1978.

[3] V. S. Borkar. *Probability theory: an advanced course.* Springer-Verlag, New York, 1995.

[4] P. W. Glynn and P. L'Ecuyer. Likelihood ratio gradient estimation for stochastic recursions. *Advances in applied probability*, 27:1019–1053, 1995.

[5] V. R. Konda and J. N. Tsitsiklis. Actor-critic algorithms. Submitted to the *SIAM Journal on Control and Optimization*, February 2001.

[6] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Springer-Verlag, 1993.

[7] E. Nummelin. A splitting technique for Harris recurrent chains. *Z. Wahrscheinlichkeitstheorie and Verw. Geb.*, 43:119–143, 1978.