Supplementary Material

A. Missing Proof of Proposition 3.2

Recall that for each iteration t, we update the ith coordinate x_i if $t \in \mathscr{T}^i$. Specifically,

$$x_i(t+1) = \begin{cases} G_i(\hat{\mathbf{x}}(t)), & t \in \mathscr{T}^i; \\ x_i(t), & t \notin \mathscr{T}^i, \end{cases} \tag{1}$$

where $\hat{\mathbf{x}}(t) := [x_1(\tau_1(t)), \ldots, x_n(\tau_n(t))]^\top$. For analysis, we sort $\mathscr{T}^i$ into a sequence $(t_k^i)_{k\geq 0}$, where t_0^i is the first element of $\mathscr{T}^i$ and t_k^i is the $(k+1)$th. Then Theorem A.1 bounds $|x_i(t) - x_i^*|$ in a staircase decreasing way: $|x_i(t) - x_i^*|$ will contract when $t \in \mathscr{T}^i$, or equivalently, $t = t_k^i$ for some k.

Theorem A.1 (Staircase Decreasing)**.** *Consider the iteration (1) under Assumption 3.1. Suppose that G is γ-contractive under infinity norm and $\mathbf{x}^*$ is the fixed-point of G. For each $t \geq B_1$ and $i \in \{1, 2, \cdots, n\}$, if $t \in (t_k^i, t_{k+1}^i]$ for some k, then $x_i(t)$ satisfies*

$$|x_i(t) - x_i^*| \leq \|\mathbf{x}(0) - \mathbf{x}^*\|_\infty \rho^{t_k^i - B_1}, \tag{2}$$

where $\rho := \gamma^{\frac{1}{B_1 + B_2 - 1}}$.

Proof. We first claim that for each $t \geq B_1$ and $i \in \{1, 2, \cdots, n\}$, there exists some $k \geq 0$ such that $t \in (t_k^i, t_{k+1}^i]$. This follows from Assumption 3.1 (a), where $t_0^i \leq B_1 - 1$, $\forall\ i$.

Now we prove Eq. (2) by induction. One could check

$$\|\mathbf{x}(t) - \mathbf{x}^*\|_\infty \leq \|\mathbf{x}(0) - \mathbf{x}^*\|_\infty,\ \forall\ t \geq 0$$

as a corollary of [1, Theorem 2] or by another induction. We skip the details here. Thus for the basic case,

$$\max_{0\leq t\leq B_1} \left\{\|\mathbf{x}(t) - \mathbf{x}^*\|_\infty \rho^{-t}\right\} \leq \max_{0\leq t\leq B_1} \left\{\|\mathbf{x}(0) - \mathbf{x}^*\|_\infty \rho^{-t}\right\} \leq \|\mathbf{x}(0) - \mathbf{x}^*\|_\infty \rho^{-B_1},$$

which gives that for each $t \leq B_1$ and $i \in \{1, 2, \cdots, n\}$,

$$|x_i(t) - x_i^*| \leq \|\mathbf{x}(0) - \mathbf{x}^*\|_\infty \rho^{t - B_1}.$$

Since ρ^t is decreasing, we can further obtain that

$$|x_i(B_1) - x_i^*| \leq \|\mathbf{x}(0) - \mathbf{x}^*\|_\infty \rho^{t_k^i - B_1},$$

if $B_1 \in (t_k^i, t_{k+1}^i]$ for some k.

For the induction step, we assume that Eq. (2) holds for all $t \geq B_1$ up to some t'. For a fixed $i \in \{1, 2, \cdots, n\}$, supposing that $t' \in (t_{k'}^i, t_{k'+1}^i]$ for some k', then we analyze the scenario at $(t'+1)$ as two cases.

Case 1: $t' \notin \mathscr{T}^i$, i.e., we do not update coordinate i at iteration t'. Hence, $x_i(t'+1) = x_i(t')$ and $t'+1 \in (t_{k'}^i, t_{k'+1}^i]$. Then Eq. (2) follows directly.

Case 2: $t' \in \mathscr{T}^i$, i.e., the ith coordinate is updated at iteration t' and $t' = t_{k'+1}^i$. Since G is γ-contractive under infinity norm, we have

$$\begin{aligned} |x_i(t'+1) - x_i^*| = |G_i(\hat{\mathbf{x}}(t)) - x_i^*| &\leq \|G(\hat{\mathbf{x}}(t)) - \mathbf{x}^*\|_\infty \\ &\leq \gamma \max_j \left|x_j(\tau_j(t')) - x_j^*\right|. \end{aligned} \tag{3}$$

For a fixed $j \in \{1, 2, \cdots, n\}$, suppose that $\tau_j(t') \in (t^j_{k_\tau}, t^j_{k_\tau+1}]$ for some k_τ. Then the induction hypothesis gives $\left|x_j(\tau_j(t')) - x^*_j\right| \le \|\mathbf{x}(0) - \mathbf{x}^*\|_\infty \rho^{t^j_{k_\tau} - B_1}$. Since $\tau_j(t') \le t^j_{k_\tau} + B_1$ by Assumption 3.1 (a) and $\tau_j(t') \ge t' - B_2 + 1$ by Assumption 3.1 (b), we obtain

$$\begin{aligned}
\gamma\left|x_j(\tau_j(t')) - x^*_j\right| &\le \gamma\|\mathbf{x}(0) - \mathbf{x}^*\|_\infty \rho^{t^j_{k_\tau} - B_1} \\
&\le \gamma\|\mathbf{x}(0) - \mathbf{x}^*\|_\infty \rho^{\tau_j(t') - 2B_1} \\
&\le \gamma\|\mathbf{x}(0) - \mathbf{x}^*\|_\infty \rho^{t' - 2B_1 - B_2 - 1} \\
&= \|\mathbf{x}(0) - \mathbf{x}^*\|_\infty \rho^{t^i_{k'+1} - B_1}, \qquad (4)
\end{aligned}$$

where the equality holds since $\gamma = \rho^{B_1+B_2-1}$ by definition and $t' = t^i_{k'+1}$. Notice that $t'+1 \in (t^i_{k'+1}, t^i_{k'+2}]$. Inserting Eq. (4) back into Eq. (3) yields the desired result. □

One may note that if $t \in (t^i_k, t^i_{k+1}]$, then $t^i_k + B_1 \ge t$ by Assumption 3.1 (a). Hence, Proposition 3.2 is a direct consequence of Theorem A.1.

B. Missing Proof from Section 4.1

To analyze the sampling error, we first review Hoeffeding's Inequality [2].

Lemma B.1 (Hoeffeding's Inequality [2])**.** *Let $X_1, \cdots, X_m$ be i.i.d real valued random variables with $X_j \in [a_j, b_j]$ and $Y = \frac{1}{m}\sum_{j=1}^m X_j$. For all $\varepsilon \ge 0$,*

$$\mathbb{P}\left[\left|Y - \mathbb{E}[Y]\right| \ge \varepsilon\right] \le 2e^{\frac{-2m^2\varepsilon^2}{\sum_{j=1}^m (b_j - a_j)^2}}.$$

By Hoeffeding's Inequality, the error between the sample averages and the true expectations can be controlled with enough number of samples. Specifically, we have:

Lemma B.2. *Given a constant L, with $K = \left\lceil \frac{8}{(1-\gamma)^4\varepsilon^2} \log(\frac{4L}{\delta}) \right\rceil$ samples, AsyncQVI returns $r(t)$ and $S(\hat{\mathbf{Q}}(t))$ satisfying*

$$|r(t) - \bar{r}^{a_t}_{i_t}| \le \frac{(1-\gamma)^2\varepsilon}{4}, \qquad |S(\hat{\mathbf{Q}}(t)) - \mathbf{p}^{a_t\top}_{i_t}\hat{\mathbf{v}}(t)| \le \frac{(1-\gamma)\varepsilon}{4}$$

with probability at least $1 - \frac{\delta}{L}$.

Proof. As we explained before, both $r(t)$ and $S(\hat{\mathbf{Q}}(t))$ are averages of K i.i.d. samples with $\mathbb{E}[r(t)] = \sum_j p^{a_t}_{i_t j} r^{a_t}_{i_t j} := \bar{r}^{a_t}_{i_t}$ and $\mathbb{E}[S(\hat{\mathbf{Q}}(t))] = \sum_j p^{a_t}_{i_t j}\hat{v}_j(t) := \mathbf{p}^{a_t\top}_{i_t}\hat{\mathbf{v}}(t)$. Since we assume $r^a_{ij} \in [0,1]$, it is easy to verify $0 \le \hat{\mathbf{v}}(t) \le \frac{1}{1-\gamma}$ by induction. We skip the details here. Then letting $K = \left\lceil \frac{8}{(1-\gamma)^4\varepsilon^2} \log\left(\frac{4L}{\delta}\right) \right\rceil$, we can obtain that

$$\begin{aligned}
&\mathbb{P}\left[|r(t) - \bar{r}^{a_t}_{i_t}| \ge \frac{(1-\gamma)^2\varepsilon}{4}\right] \le 2e^{\frac{-2K^2(1-\gamma)^4\varepsilon^2}{16K}} \le \frac{\delta}{2L}; \\
&\mathbb{P}\left[|S(\hat{\mathbf{Q}}(t)) - \mathbf{p}^{a_t\top}_{i_t}\hat{\mathbf{v}}(t)| \ge \frac{(1-\gamma)\varepsilon}{4}\right] \le 2e^{\frac{-2K^2(1-\gamma)^4\varepsilon^2}{16K}} \le \frac{\delta}{2L}.
\end{aligned}$$

□

Proof. [Proof of Proposition 4.3] For a fixed iteration t, by Lemma B.2,

$$\left|r(t) + \gamma S(\hat{\mathbf{Q}}(t)) - \bar{r}^{a_t}_{i_t} - \gamma\mathbf{p}^{a_t\top}_{i_t}\hat{\mathbf{v}}(t)\right| \le |r(t) - \bar{r}^{a_t}_{i_t}| + \gamma|S(\hat{\mathbf{Q}}(t)) - \mathbf{p}^{a_t\top}_{i_t}\hat{\mathbf{v}}(t)| \le \frac{(1-\gamma)\varepsilon}{4}$$

holds with probability at least $1-\frac{\delta}{L}$. Taking a union bound over all $0 \le t \le L-1$ iterations gives the desired result. $\square$

Proof. [Proof of Proposition 4.4] We denote by $\mathcal{E}_1$ the event

$$\Big\{\big|r(t)+\gamma S(\hat{\mathbf{Q}}(t))-\bar{r}_{i_t}^{a_t}-\gamma \mathbf{p}_{i_t}^{a_t\top}\hat{\mathbf{v}}(t)\big| \le \frac{(1-\gamma)\varepsilon}{4},\ \forall\ 0\le t\le L-1\Big\}.$$

By Proposition 4.3, $\mathcal{E}_1$ occurs with probability at least $1-\delta$. Next, we condition on $\mathcal{E}_1$ and prove

$$\|\mathbf{Q}(t)-\mathbf{Q}^{\mathbb{E}}(t)\|_\infty \le \frac{\varepsilon}{2},\ \forall\ 1\le t\le L \tag{5}$$

by induction. The basic case is trivial. For the induction step, we analyze the scenario at $t+1$ as two cases. When $t\notin\mathscr{T}^{i,a}$, $|Q_{i,a}(t+1)-Q^{\mathbb{E}}_{i,a}(t+1)|\le\varepsilon/2$ follows from the hypothesis, since Eqs. (5) and (6) give that

$$Q_{i,a}(t+1)-Q^{\mathbb{E}}_{i,a}(t+1)=Q_{i,a}(t)-Q^{\mathbb{E}}_{i,a}(t).$$

When $t\in\mathscr{T}^{i,a}$, by Eq. (5), Eq. (6) and triangle inequality, we have that

$$\begin{aligned}
&\big|Q_{i,a}(t+1)-Q^{\mathbb{E}}_{i,a}(t+1)\big|\\
=&\left|r(t)+\gamma S(\hat{\mathbf{Q}}(t))-\frac{(1-\gamma)\varepsilon}{4}-\bar{r}_i^a-\gamma\sum_j p_{ij}^a\max_{a'}\hat{Q}^{\mathbb{E}}_{j,a'}(t)\right|\\
\le&\left|r(t)+\gamma S(\hat{\mathbf{Q}}(t))-\bar{r}_i^a-\gamma\mathbf{p}_i^{a\top}\hat{\mathbf{v}}(t)-\frac{(1-\gamma)\varepsilon}{4}\right|+\left|\gamma\mathbf{p}_i^{a\top}\hat{\mathbf{v}}(t)-\gamma\sum_j p_{ij}^a\max_{a'}\hat{Q}^{\mathbb{E}}_{j,a'}(t)\right|\\
\le&\left|r(t)+\gamma S(\hat{\mathbf{Q}}(t))-\bar{r}_i^a-\gamma\mathbf{p}_i^{a\top}\hat{\mathbf{v}}(t)\right|+\frac{(1-\gamma)\varepsilon}{4}+\gamma\sum_j p_{ij}^a\big|\max_{a'}\hat{Q}_{j,a'}(t)-\max_{a'}\hat{Q}^{\mathbb{E}}_{j,a'}(t)\big|.
\end{aligned}$$

By definition of $\mathcal{E}_1$ and the induction hypothesis, we further obtain that

$$|Q_{i,a}(t+1)-Q^{\mathbb{E}}_{i,a}(t+1)|\le\frac{(1-\gamma)\varepsilon}{4}+\frac{(1-\gamma)\varepsilon}{4}+\gamma\frac{\varepsilon}{2}=\frac{\varepsilon}{2},$$

which completes the proof. $\square$

Proof. [Proof of Theorem 4.5] By Proposition 3.2,

$$\|\mathbf{Q}^*-\mathbf{Q}^{\mathbb{E}}(L)\|_\infty\le(1-\gamma)^{-1}\rho^{L-2B_1}=(1-\gamma)^{-1}\gamma^{\frac{L-2B_1}{B_1+B_2-1}}.$$

Notice that $\gamma=(1-(1-\gamma))\le e^{-(1-\gamma)}$. We have that

$$\|\mathbf{Q}^*-\mathbf{Q}^{\mathbb{E}}(L)\|_\infty\le(1-\gamma)^{-1}e^{-(1-\gamma)\frac{L-2B_1}{B_1+B_2-1}}\le\frac{\varepsilon}{2}, \tag{6}$$

where the last inequality holds with $L=\left\lceil 2B_1+\frac{B_1+B_2-1}{1-\gamma}\log\left(\frac{2}{(1-\gamma)\varepsilon}\right)\right\rceil$. Then, by Proposition 4.4, with probability at least $1-\delta$,

$$\|\mathbf{Q}^{\mathbb{E}}(L)-\mathbf{Q}(L)\|_\infty\le\frac{\varepsilon}{2}. \tag{7}$$

Inserting Eq. (7) back into Eq. (6) gives the desired result

$$\|\mathbf{Q}^*-\mathbf{Q}(L)\|_\infty\le\|\mathbf{Q}^*-\mathbf{Q}^{\mathbb{E}}(L)\|_\infty+\|\mathbf{Q}^{\mathbb{E}}(L)-\mathbf{Q}(L)\|_\infty\le\varepsilon.$$

Then one can check $\|\mathbf{v}^*-\mathbf{v}(L)\|_\infty\le\varepsilon$ at ease. $\square$

C. Missing Proof of Theorem 4.6

After L iterations, AsyncQVI returns a policy $\pi(L)$ with $\pi_i(L) = \arg\max_{a\in\mathcal{A}} Q_{i,a}(L)$. To show that $\pi(L)$ is ε-optimal, we first define a policy operator.

Definition C.1 (Policy Operator)**.** *Given a policy π and a vector $\mathbf{v} \in \mathbb{R}^{|\mathcal{S}|}$, the policy operator T_π: $\mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^{|\mathcal{S}|}$ is defined as*

$$[T_\pi \mathbf{v}]_i = \bar{r}_i^{\pi_i} + \gamma \mathbf{p}_i^{\pi_i\top}\mathbf{v} = r_i^{\pi_i} + \gamma \sum_{j\in\mathcal{S}} p_{ij}^{\pi_i} v_j. \tag{8}$$

Proposition C.2 (T_π's Properties)**.** *Given a policy π, for any vectors $\mathbf{v}$, $\mathbf{v}' \in \mathbb{R}^{\mathcal{S}}$,*

(a) ***Monotonicity:*** *if $\mathbf{v} \le \mathbf{v}'$, then $T_\pi \mathbf{v} \le T_\pi \mathbf{v}'$.*
(b) ***γ-Contraction:*** *$\|T_\pi \mathbf{v} - T_\pi \mathbf{v}'\|_\infty \le \gamma \|\mathbf{v} - \mathbf{v}'\|_\infty$.*
(c) *$\mathbf{v}^\pi$ is the fixed-point of T_π.*

The proof is straightforward following the definition. We skip the details here.

Lemma C.3. [3] *Given a policy π, for any vector $\mathbf{v} \in \mathbb{R}^{\mathcal{S}}$, if there exists a $\mathbf{v}' \in \mathbb{R}^{\mathcal{S}}$ such that $\mathbf{v}' \le \mathbf{v}$ and $\mathbf{v} \le T_\pi \mathbf{v}'$, then $\mathbf{v} \le \mathbf{v}^\pi$.*

Proof. By Proposition C.2 (a) and $\mathbf{v}' \le \mathbf{v}$, we first have $T_\pi \mathbf{v}' \le T_\pi \mathbf{v}$. Combining with $\mathbf{v} \le T_\pi \mathbf{v}'$, we further obtain $\mathbf{v} \le T_\pi \mathbf{v}$. By induction, one can check $\mathbf{v} \le T_\pi^n \mathbf{v}$, $\forall\ n \in \mathbb{N}$. Moreover, since T_π is a γ-contraction, $\mathbf{v}^\pi = \lim_{n\to\infty} T_\pi^n \mathbf{v}$. Hence, $\mathbf{v} \le \lim_{n\to\infty} T_\pi^n \mathbf{v} = \mathbf{v}^\pi$. □

Next, we consider the special case that $\mathbf{v}(L)$ and $\pi(L)$ are both derived from AsyncQVI with

$$\pi_i(L) = \arg\max_a Q_{i,a}(L), \quad v_i(L) = \max_a Q_{i,a}(L),\ \forall i \in \mathcal{S}.$$

If $\|\mathbf{v}^* - \mathbf{v}^\pi\|_\infty \le \varepsilon$, then π is $\varepsilon-$optimal. To achieve this, we first show that $\mathbf{v}(L)$ satisfies Lemma C.3 (see Lemma C.4). Then with Theorem 4.5, $\|\mathbf{v}^* - \mathbf{v}^\pi\|_\infty \le \|\mathbf{v}^* - \mathbf{v}(L)\|_\infty \le \varepsilon$.

Lemma C.4. *Under Assumption 3.1, AsyncQVI generates a sequence of $\{\mathbf{v}(t)\}_{t=1}^L$ and $\{\pi(t)\}_{t=1}^L$ satisfying*

$$\mathbf{v}(t-1) \le \mathbf{v}(t) \le T_{\pi(t)}\mathbf{v}(t-1),\ \forall\ 1 \le t \le L \tag{9}$$

with probability at least $1-\delta$.

Proof. By Proposition 4.3,

$$\left|r(t) + \gamma S(\hat{\mathbf{Q}}(t)) - \bar{r}_{i_t}^{a_t} - \gamma \mathbf{p}_{i_t}^{a_t\top}\hat{\mathbf{v}}(t)\right| \le \frac{(1-\gamma)\varepsilon}{4},\ \forall\ 0 \le t \le L-1$$

holds with probability at least $1-\delta$. Denote by $\mathcal{E}_2$ the event

$$\left\{r(t) + \gamma S(\hat{\mathbf{Q}}(t)) - \frac{(1-\gamma)\varepsilon}{4} \le \bar{r}_{i_t}^{a_t} + \gamma \mathbf{p}_{i_t}^{a_t\top}\hat{\mathbf{v}}(t),\ \forall\ 0 \le t \le L-1\right\}.$$

Then $\mathcal{E}_2$ occurs with probability at least $1-\delta$.

Now we condition on $\mathcal{E}_2$ and prove Eq. (9) by induction. For simplicity, we let $\mathbf{v}(-1) = \mathbf{v}(0) = \mathbf{0}$ and start our proof from $t = 0$. Then the basic case holds. For the induction step, suppose that Eq. (9) is true for all t up to some t'. Recall that in AsyncQVI, for each iteration, whether v_i or π_i will be updates depends on the value of $Q_{i,a}$. We hence analyze the scenario at $(t'+1)$ as two cases.

Case 1: $Q_{i_{t'},a_{t'}}(t'+1) \leq v_{i_{t'}}(t')$. Then $\mathbf{v}$ and π will not be updated, i.e., $\mathbf{v}(t'+1) = \mathbf{v}(t')$ and $\pi(t'+1) = \pi(t')$. In this case, the inequality $\mathbf{v}(t') \leq \mathbf{v}(t'+1)$ follows directly. For the other part, by induction hypothesis we have

$$\mathbf{v}(t'+1) = \mathbf{v}(t') \leq T_{\pi(t')}\mathbf{v}(t'-1) = T_{\pi(t'+1)}\mathbf{v}(t'-1) \leq T_{\pi(t'+1)}\mathbf{v}(t'),$$

where the last inequality comes from $\mathbf{v}(t'-1) \leq \mathbf{v}(t')$ and the monotonicity of $T_{\pi(t'+1)}$.
Case 2: $Q_{i_{t'},a_{t'}}(t'+1) > v_{i_{t'}}(t')$. Then $\forall\ i \in \mathcal{S}$,

Case 2.1: $i \neq i_{t'}$. In this case, $v_i(t'+1) = v_i(t')$ and $\pi_i(t'+1) = \pi_i(t')$. Hence, once again by induction hypothesis and T_π's monotonicity, we obtain

$$v_i(t'+1) = v_i(t') \leq \left[T_{\pi(t')}\mathbf{v}(t'-1)\right]_i = \left[T_{\pi(t'+1)}\mathbf{v}(t'-1)\right]_i \leq \left[T_{\pi(t'+1)}\mathbf{v}(t')\right]_i.$$

Case 2.2: $i = i_{t'}$. According to Lines 8 and 10 of Algorithm 2, the ith coordinate of $\mathbf{v}$ is updated at iteration t' and the former inequality follows directly. For the latter inequality, by Line 7 of Algorithm 2 we have

$$v_i(t'+1) = Q_{i,a_{t'}}(t'+1) = r(t') + \gamma S(\hat{\mathbf{Q}}(t')) - \frac{(1-\gamma)\varepsilon}{4}.$$

By definition of $\mathcal{E}_2$ and $\pi_i(t'+1) = a_{t'}$, we obtain

$$v_i(t'+1) \leq \bar{r}_i^{a_{t'}} + \gamma \mathbf{p}_i^{a_{t'}\top}\hat{\mathbf{v}}(t') = \left[T_{\pi(t'+1)}\hat{\mathbf{v}}(t')\right]_i.$$

Owing to $\hat{\mathbf{v}}(t') \leq \mathbf{v}(t')$ by induction hypothesis and the monotonicity of $T_{\pi(t'+1)}$, we can complete our proof by

$$v_i(t'+1) \leq \left[T_{\pi(t'+1)}\hat{\mathbf{v}}(t')\right]_i \leq \left[T_{\pi(t'+1)}\mathbf{v}(t')\right]_i.$$

□

Finally, combining the results of Lemma C.3, Lemma C.4 and Theorem 4.5, we can establish Theorem 4.6 at ease.

References

[1] Hamid Reza Feyzmahdavian and Mikael Johansson. On the convergence rates of asynchronous iterations. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 153–159. IEEE, 2014.

[2] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[3] Aaron Sidford, Mengdi Wang, Xian Wu, and Yinyu Ye. Variance reduced value iteration and faster algorithms for solving markov decision processes. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 770–787. Society for Industrial and Applied Mathematics, 2018.