Corrigendum to "Value Set Iteration for Markov Decision Processes" [Automatica 50 (2014) 1940–1943]

Hyeong Soo Chang

Abstract

This corrigendum corrects the proof of the part (2), Theorem 2, in the paper, "value set iteration for Markov decision processes," *Automatica*, Vol. 50, pp. 1940–1943, 2014.

I. Introduction

This corrigendum corrects the problems in the proof of Theorem 2, the part (2) in the paper, "value set iteration for Markov decision processes," [1]. All notations are the same as those used in [1].

II. Correction

The following is the corrected statement of the part (2), Theorem 2: For the sequence $\{V_k\}$ generated by VSI and the policy π_k defined such that for all $x \in X$,

$$\pi_k(x) \in \arg\max_{a \in A(x)} \Big(R(x,a) + \gamma \sum_{y \in X} P^a_{xy} V_{k+1}(y) \Big),$$

we have that for any $\{\Delta_k\}$, if $||V_{k+1} - V_k|| \leq \epsilon \cdot \frac{1-\gamma}{2\gamma}$, then $||V^* - V^{\pi_k}|| \leq \frac{\epsilon}{1-\gamma}$.

That is, the policy π_k is $\frac{\epsilon}{1-\gamma}$-optimal policy. Note that if we change the stopping condition of $||V_{k+1} - V_k|| \leq \epsilon \cdot \frac{1-\gamma}{2\gamma}$ in the VSI description [1] to $||V_{k+1} - V_k|| \leq \epsilon \cdot \frac{(1-\gamma)^2}{2\gamma}$, π_k is ϵ-optimal policy.

Proof: We have that $||V^* - V^{\pi_k}|| \leq ||V^* - V_{k+1}|| + ||V_{k+1} - V^{\pi_k}||$.

For the first term of $||V^* - V_{k+1}||$ in the inequality,

$$\begin{aligned} ||V^* - V_{k+1}|| &\leq ||V^* - T(V_{k+1}, \Delta_k)|| + ||T(V_{k+1}, \Delta_k) - V_{k+1}|| \\ &= ||T(V^*, \Delta_k) - T(V_{k+1}, \Delta_k)|| + ||T(V_{k+1}, \Delta_k) - T(V_k, \Delta_k)|| \\ &\leq \gamma||V^* - V_{k+1}|| + \gamma||V_{k+1} - V_k|| \text{ by Lemma 1}. \end{aligned}$$

Therefore,

$$||V^* - V_{k+1}|| \leq \frac{\gamma}{1-\gamma}||V_{k+1} - V_k||.$$

For the second term,

$$\begin{aligned} ||V^{\pi_k} - V_{k+1}|| &\leq ||L_{\pi_k}(V^{\pi_k}) - L(V_{k+1})|| + ||L(V_{k+1}) - V_{k+1}|| \\ &= ||L_{\pi_k}(V^{\pi_k}) - L_{\pi_k}(V_{k+1})|| + ||L(V_{k+1}) - V_{k+1}|| \\ &\leq \gamma||V^{\pi_k} - V_{k+1}|| + ||L(V_{k+1}) - V^*|| + ||V^* - V_{k+1}|| \\ &= \gamma||V^{\pi_k} - V_{k+1}|| + ||L(V_{k+1}) - L(V^*)|| + ||V^* - V_{k+1}|| \\ &\leq \gamma||V^{\pi_k} - V_{k+1}|| + \gamma||V_{k+1} - V^*|| + ||V^* - V_{k+1}|| \\ &= \gamma||V^{\pi_k} - V_{k+1}|| + (1+\gamma)||V^* - V_{k+1}||. \end{aligned}$$

Therefore, we have that

$$||V^{\pi_k} - V_{k+1}|| \leq \frac{1+\gamma}{1-\gamma} \cdot ||V^* - V_{k+1}|| \leq \frac{\gamma(1+\gamma)}{(1-\gamma)^2} \cdot ||V_{k+1} - V_k||.$$

It follows that

$$||V^* - V^{\pi_k}|| \leq \frac{2\gamma}{(1-\gamma)^2}||V_{k+1} - V_k|| \leq \frac{\epsilon}{1-\gamma}.$$

■

References

[1] H. S. Chang, "Value set iteration for Markov decision processes," *Automatica*, vol. 50, pp. 1940–1943, 2014.