Accelerating Value Iteration with Anchoring

Jongmin Lee¹, Ernest K. Ryu ^{1,2}

Neural Information Processing System 2023

¹Department of Mathematical Sciences, Seoul National University ²Interdisciplinary Program in Artificial Intelligence, Seoul National University

Value Iteration

With the given Markov Decision Process $(\mathcal{S},\mathcal{A},P,r,\gamma)$ and policy $\pi,$ define Bellman operators as

$$\begin{split} T^{\pi}V(s) &= \mathbb{E}_{a \sim \pi(\cdot \mid s), s' \sim P(\cdot \mid s, a)} \left[r(s, a) + \gamma V(s') \right], \\ T^{\star}V(s) &= \sup_{a \in \mathcal{A}} \left\{ r(s, a) + \gamma \mathbb{E}_{s' \sim P(\cdot \mid s, a)} \left[V(s') \right] \right\}. \end{split}$$

Value Iteration (VI) is a fixed-point iteration with $T(=T^{\pi} \text{ or } T^{\star})$:

$$V^{k+1} = TV^k. \tag{VI}$$

For $0 < \gamma < 1$, VI exhibits the rate

$$\left\|TV^{k}-V^{k}\right\|_{\infty}\leq(1+\gamma)\gamma^{k}\left\|V^{0}-V^{\star}\right\|_{\infty}$$

Question) If $\gamma \approx 1$, VI has rate $\mathcal{O}(1)$. Can we accelerate this in terms of Bellman error?

Anchored Value Iteration

Anchored Value Iteration (Anc-VI):

$$V^k = \beta_k V^0 + (1 - \beta_k) T V^{k-1}$$

where $\beta_k = 1/(\sum_{i=0}^k \gamma^{-2i}).$

We call the $\beta_k V_0$ term the *anchor term* since it serves to pull the iterates toward the starting point V_0 . The strength of the anchor mechanism diminishes since β_k is a decreasing sequence.

Accelerated rate of Anc-VI

Theorem For $0 < \gamma < 1$, if $V^0 \le TV^0$ or $V^0 \ge TV^0$, Anc-VI exhibits the rate

$$\begin{aligned} \|TV^{k} - V^{k}\|_{\infty} &\leq \frac{\left(\gamma^{-1} - \gamma\right)\left(1 + \gamma - \gamma^{k+1}\right)}{\left(\gamma^{k+1}\right)^{-1} - \gamma^{k+1}} \|V^{0} - V^{\star}\|_{\infty} \\ &= \left(\frac{1}{k+1} + \frac{k}{k+1}\epsilon + O(\epsilon^{2})\right)\|V^{0} - V^{\star}\|_{\infty} \end{aligned}$$

where $\epsilon = 1 - \gamma$ and the big- \mathcal{O} notation considers the limit $\epsilon \to 0$.

This implies that

$$\underbrace{\frac{\left(\gamma^{-1}-\gamma\right)\left(1+\gamma-\gamma^{k+1}\right)}{\left(\gamma^{k+1}\right)^{-1}-\gamma^{k+1}}}_{\text{rate of Anc-VI}} < \underbrace{\left(1+\gamma\right)\gamma^{k}}_{\text{rate of VI}} \qquad \text{for } 0 < \gamma < 1.$$

Complexity lower bound

Theorem

Let $k \ge 0$, $n \ge k+2$, $0 < \gamma \le 1$, and $V^0 \in \mathbb{R}^n$. Then there exists an MDP with $|\mathcal{S}| = n$ and $|\mathcal{A}| = 1$ (which implies the Bellman consistency and optimality operator all coincide as $T : \mathbb{R}^n \to \mathbb{R}^n$) such that T has a fixed point V^* satisfying $V^0 \le V^*$ and

$$\left\|TV^{k} - V^{k}\right\|_{\infty} \geq \frac{\gamma^{k}}{\sum_{i=0}^{k} \gamma^{i}}^{3} \left\|V^{0} - V^{\star}\right\|_{\infty}$$

for any iterates $\{V^i\}_{i=0}^k$ satisfying

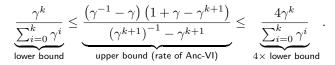
$$V^{i} \in V^{0} + \operatorname{span}\{TV^{0} - V^{0}, TV^{1} - V^{1}, \dots, TV^{i-1} - V^{i-1}\}$$

for i = 1, ..., k.

 $^3 \text{Our}$ Theorem improves upon the prior state-of-the-art complexity lower bound established in the (V. Goyal , 2020) by a factor $1-\gamma^{k+1}$. V. Goyal and J. Grand-Clément. A first-order approach to accelerated value iteration. Opera- tions Research, 2022.

Optimality of Anc-VI

By previous two theorems,



This implies that Anc-VI is optimal up to a constant of factor 4 while VI is suboptimal.



Anc-VI reduces the Bellman error faster than VI. Furthermore, Anc-VI is optimal method up to factor of 4.

We also show that the anchoring mechanism provides the same benefit in the approximate VI and Gauss–Seidel VI setups as well.