Accelerating Value Iteration with Anchoring

Jongmin Lee[1] , Ernest K. Ryu [1,2]

Neural Information Processing System 2023

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

Value Iteration

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

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

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

$$V^{k+1} = TV^{k}. \qquad \text{(VI)}$$

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 $V^0 \leq TV^0$ or $V^0 \geq TV^0$, Anc-VI exhibits the rate

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

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

This implies that

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

Complextiy lower bound

Theorem

Let $k \geq 0$, $n \geq k+2$, $0 < \gamma \leq 1$, 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 $T\colon \mathbb{R}^n \to \mathbb{R}^n$) such that T has a fixed point $V^\star$ satisfying $V^0 \leq V^\star$ 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

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

for $i = 1, \ldots, k$.

[3]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,

$$\underbrace{\frac{\gamma^k}{\sum_{i=0}^k \gamma^i}}_{\text{lower bound}} \le \underbrace{\frac{\left(\gamma^{-1}-\gamma\right)\left(1+\gamma-\gamma^{k+1}\right)}{\left(\gamma^{k+1}\right)^{-1}-\gamma^{k+1}}}_{\text{upper bound (rate of Anc-VI)}} \le \underbrace{\frac{4\gamma^k}{\sum_{i=0}^k \gamma^i}}_{4\times\text{ lower bound}}.$$

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

Summary

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.