Class 12

Shikhar Saxena

February 17, 2023

Contents

Recap

$$V_\alpha(s) = \max_a \left\{ r(s,a) + \alpha \sum P(j|s,a) V_\alpha(j) \right\} \tag{1}$$

$$V_f(s) = r(s, f(s)) + \alpha \sum P(j|s, f(s)) V_f(j) \tag{2}$$

$$T_\alpha u(s) = \max_a \left\{ r(s,a) + \alpha \sum P(j|s,a) u(j) \right\} \tag{3}$$

$$T_f u(s) = r(s, f(s)) + \alpha \sum P(j|s, f(s)) u(j) \tag{4}$$

Essentially, Policy Iteration can be approximated to Value Iteration algorithm by

$$\lim_{n \to \infty} T_\alpha^n u = V_\alpha$$

Algorithm 1: Value Iteration

1. Start with an arbitrary initial vector $u \in B(\mathcal{S})$ and set $n = 0$.
2. For each s find $V_{n+1}(s)$ using (1).
3. If $\|V_{n+1} - V_n\| \leq \epsilon$ for all states then stop. Else repeat for the next n.
4. Then get policy using argmax.

Bounds on $\|V_n - V_\alpha\|$

We already know $\|V_n - V_\alpha\| \leq \alpha^n \|V_0 - V_\alpha\|$ but this is not useful.

Similarly, $\|V_n - V_{n+1}\| \leq \alpha^n \|V_0 - V_1\|$

So we want to obtain a bound on $\|V_n - V_\alpha\|$ in terms of $\|V_0 - V_1\|$. We'll see how this helps.

Using Triangle Inequality,

$$
\begin{aligned}
\|V_n - V_\alpha\| &= \|(V_n - V_{n+1}) + (V_{n+1} - V_{n+2}) + \dots (V_{n+l} - V_\alpha)\| \\
&\leq \alpha^n \|V_1 - V_0\| (1 + \alpha + \alpha^2 \dots + \alpha^{l-1}) + \|V_{n+l} - V_\alpha\| \\
&\leq \frac{\alpha^n}{1-\alpha} \|V_1 - V_0\| \quad \text{Setting } l \to \infty
\end{aligned}
$$

Stopping Criteria

$$
\|V_n - V_\alpha\| \leq \frac{\alpha}{1-\alpha} \|V_n - V_{n-1}\|
$$

Proof.

$$
\begin{aligned}
\|V_n - V_\alpha\| &\leq \|V_n - V_{n+1}\| + \|V_{n+1} - V_\alpha\| \\
&\leq \alpha \|V_{n-1} - V_n\| + \alpha \|V_n - V_\alpha\|
\end{aligned}
$$

□

Now, assume we want to stop at a δ where $\|V_n - V_\alpha\| \leq \delta$.

That means $\frac{\alpha}{1-\alpha} \|V_n - V_{n-1}\| \leq \frac{\alpha\epsilon}{1-\alpha}$ which gives us $\delta = \frac{\alpha\epsilon}{1-\alpha}$.

Gaus-Seidel or In-Place (Asynchronous) Value Iteration

Don't keep separate vectors for V_n and V_{n+1}. Just solve them in a single vector V.

Modified Policy Iteration

Taking something from both Policy and Value Iteration. VI is faster per iteration than PI but PI takes less iterations.

Essentially don't evaluate V_{π_n} fully for an intermediate policy π_n.

When $m_n = 1$ then this algorithm gives us VI.

Algorithm 2: Modified PI

1. Set $n = 0$ and arbitrary V_0 and find π_0 that is greedy wrt V_0
2. (Partial Policy Evaluation): Obtain V_n by repeatedly appling T_{π_n} on V_{n-1} for m_n number of times.
3. (Greedy Step): Find policy π^{n+1} that is greedy on V_n.
4. If $\|V_{n+1} - V_n\| \leq \epsilon$ then STOP.