

## 10 Value Iteration

### 10.1 Jacobi Value Iteration

Value iteration is perhaps the most widely used algorithm in dynamic programming because it is the simplest to implement. It is virtually identical to backward dynamic programming for finite horizon problems. There are several variants of value iteration. The basic version of the value iteration algorithm is given below.

**Step 0** Initialization:

- Set  $V_0(s) = 0 \quad \forall s \in \mathcal{S}$ .
- Fix a tolerance parameter  $\epsilon > 0$ .
- Set  $i = 0$  (iteration counter).

**Step 1** For each  $s \in \mathcal{S}$  compute:

$$V_{i+1}(s) = \max_{a \in \mathcal{A}} \left[ c(s, a) + \alpha \sum_{y \in \mathcal{S}} \mathbb{P}[y|s, a] V_i(y) \right] \quad (1)$$

or

$$V_{i+1} = T(V_i)$$

**Step 2** If  $\|V_{i+1} - V_i\|_\infty < \frac{1-\alpha}{2\alpha} \epsilon$ , then stop/go to step 3. Otherwise, let  $i \leftarrow i + 1$ , go to step 1.

**Step 3** Let

$$a_\epsilon(s) \in \arg \max_{a \in \mathcal{A}(s)} \left\{ c(s, a) + \alpha \sum_{y \in \mathcal{S}} p[y|s, a] V_{i+1}(y) \right\} \quad (2)$$

It is easy to see that the value iteration algorithm is similar to the backward dynamic programming algorithm. Rather than using a subscript  $t$ , which we decrement from  $T$  back to 0, we use an iteration counter  $i$  that starts at 0 and increases until we satisfy a convergence criterion. Here, we stop the algorithm when

$$\|V_{i+1} - V_i\|_\infty < \epsilon(1 - \alpha)/2\alpha$$

Thus, we stop if the largest change in the value of being in any state is less than  $\epsilon(1 - \alpha)/2\alpha$  where  $\epsilon$  is a specified error tolerance.

**Theorem 10.1** *let  $V_0 \in \mathcal{V}$ ,  $\epsilon > 0$ ,  $V_i = T^i(V_0)$  as in the value iteration algorithm, then*

1.  $V_i \rightarrow V^*$  uniformly (in  $\|\cdot\|_\infty$  - norm) as  $i \rightarrow \infty$ .
2. There exists a finite  $N$  such that

$$\|V_{i+1} - V_i\|_\infty < \frac{1 - \alpha}{2\alpha} \epsilon \quad \forall i \geq N$$

3. The stationary deterministic policy  $\Pi_\epsilon(s) \triangleq a_\epsilon(s)$  is  $\epsilon$ -optimal, that is,

$$\|V^{\Pi_\epsilon} - V^*\|_\infty < \epsilon$$

4.  $\|V_{i+1} - V^*\|_\infty < \frac{\epsilon}{2}$  whenever  $\|V_{i+1} - V_i\|_\infty < \frac{1 - \alpha}{2\alpha} \epsilon$